GenCore version 4.5

OM nucleic - nucleic search, using sw model (Smith-Waterman)

Run on: February 7, 2002, 08:21:14 ; Search time 3842.15 Seconds
(without alignments)
1807.663 Million cell updates/sec

Title: US-09-394-745-5893
Perfect score: 421
Sequence: 1 gaaaaaaataactcggaaaa.....ccatgttggttcctgcatgc 421

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_om:*
20: em_or:*
21: em_ov:*
22: em_pat:*
23: em_ph:*
24: em_pl:*
25: em_ro:*
26: em_sts:*
27: em_sy:*

28: em_un:*
29: em_vi:*
30: em_htgo_hum:*
31: em_htgo_inv:*
32: em_htgo_rod:*
33: em_htg_hum:*
34: em_htg_inv:*
35: em_htg_rod:*
36: em_htg_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	59.8	14.2	91924	8	ATAC009606	AC009606 Arabidops
c	2	59.8	14.2	100887	8	ATAC011620	AC011620 Arabidops
	3	56	13.3	255	11	G69921	G69921 695251131FN
	4	56	13.3	387	11	G69610	G69610 695251131FB
	5	54.4	12.9	368	11	G70259	G70259 695251131FB
c	6	52	12.4	27588	8	AF180335	AF180335 Glycine m
	7	51.2	12.2	93672	8	AC069556	AC069556 Genomic S
c	8	49.6	11.8	7218	6	I66494	I66494 Sequence 14
c	9	45.6	10.8	26703	3	CBRG45O11	AC084652 Caenorhab
	10	43.4	10.3	235411	2	AC084066	AC084066 Mus muscu
	11	42.8	10.2	187278	2	AC067840	AC067840 Homo sapi
c	12	42.4	10.1	221104	2	AC025579	AC025579 Mus muscu
c	13	42.4	10.1	241714	2	AC083834	AC083834 Mus muscu
	14	42	10.0	143977	2	AC092197	AC092197 Homo sapi
c	15	41.8	9.9	84199	2	AC007426	AC007426 Homo sapi
	16	41.8	9.9	173356	2	AC024360	AC024360 Homo sapi
c	17	41.6	9.9	2275	10	MDCS45GE	X60026 M.domesticu
	18	41.6	9.9	158696	9	AF238380	AF238380 Homo sapi
	19	41.6	9.9	171901	2	AC027240	AC027240 Homo sapi
	20	41.4	9.8	211617	2	AC063958	AC063958 Homo sapi
	21	41.2	9.8	44274	1	SCH63	AL442629 Streptomy
	22	41	9.7	185548	10	AC005743	AC005743 Mus muscu
	23	41	9.7	230372	2	AC073693	AC073693 Mus muscu
c	24	40.2	9.5	80772	2	AC023012	AC023012 Homo sapi
c	25	40.2	9.5	197022	2	AL513346	AL513346 Mus muscu
	26	40	9.5	205271	2	AC035247	AC035247 Mus muscu
c	27	39.8	9.5	166164	2	AL512601	AL512601 Homo sapi
c	28	39.6	9.4	105574	2	AC013308	AC013308 Homo sapi
c	29	39.4	9.4	168580	2	AL391994	AL391994 Homo sapi
	30	39.4	9.4	175072	2	AC022358	AC022358 Homo sapi
	31	39.4	9.4	265985	2	AC087226	AC087226 Mus muscu
c	32	39.2	9.3	211393	2	AL513347	AL513347 Mus muscu
	33	39	9.3	166420	2	AC012112	AC012112 Homo sapi
	34	39	9.3	166642	2	AC009543	AC009543 Homo sapi
	35	39	9.3	194853	2	AL591125	AL591125 Mus muscu
	36	39	9.3	199956	3	CNS06C8G	AL391737 chromosom
c	37	39	9.3	199956	3	CNS06C8G	AL391737 chromosom
	38	39	9.3	218476	10	AC087116	AC087116 Mus Muscu

c	39	39	9.3	237486	2	AC022780	AC022780 Mus muscu
c	40	38.8	9.2	119418	2	AC079834	AC079834 Homo sapi
c	41	38.8	9.2	196361	9	AC005386	AC005386 citb_57_1
	42	38.8	9.2	208330	2	AC022327	AC022327 Mus muscu
c	43	38.8	9.2	222775	2	AC074310	AC074310 Mus muscu
c	44	38.6	9.2	204653	10	AC005302	AC005302 Mus muscu
c	45	38.4	9.1	147765	2	AC084060	AC084060 Mus muscu

ALIGNMENTS

RESULT 1
ATAC009606/c
LOCUS ATAC009606 91924 bp DNA PLN 25-JAN-2001
DEFINITION Arabidopsis thaliana chromosome III BAC F22F7 genomic sequence, complete sequence.
ACCESSION AC009606
VERSION AC009606.4 GI:12484386
KEYWORDS HTG.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 91924)
AUTHORS Lin,X., Kaul,S., Town,C.D., Benito,M., Creasy,T.H., Haas,B., Wu,D., Maiti,R., Ronning,C.M., Koo,H., Fujii,C.Y., Utterback,T.R., Barnstead,M.E., Bowman,C.L., White,O., Nierman,W.C. and Fraser,C.M.
TITLE Arabidopsis thaliana chromosome III BAC F22F7 genomic sequence
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 91924)
AUTHORS Lin,X. and Kaul,S.
TITLE Direct Submission
JOURNAL Submitted (28-AUG-1999) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA, xlin@tigr.org
REFERENCE 3 (bases 1 to 91924)
AUTHORS Lin,X.
TITLE Direct Submission
JOURNAL Submitted (25-JAN-2001) The Institute for Genomic Research, 9712 Medical Center Dr., Rockville, MD 20850, USA
COMMENT On Jan 25, 2001 this sequence version replaced gi:12280792.
Address all correspondence to:at@tigr.org

BAC clone F22F7 is from Arabidopsis chromosome III and is near the molecular marker mi172.
The orientation of the sequence is from SP6 to T7 end of the BAC clone.
Genes were identified by a combination of three methods: Gene prediction programs including GRAIL (available by anonymous ftp from arthur.epm.ornl.gov), Genefinder (Phil Green, University of Washington), Genscan (Chris Burge, http://gnomic.stanford.edu/~chris/GENSCANW.html), and NetPlantGene (http://www.cbs.dtu.dk/netpgene/cbsnetpgene.html), searches of the complete sequence against a peptide database and the Arabidopsis EST database at TIGR (http://www.tigr.org/tdb/at/at.html).
Annotated genes are named to indicate the level of evidence for

```
            their annotation. Genes with similarity to other proteins are named
            after the database hits. Genes without significant peptide
            similarity but with EST similarity are named as 'unknown' proteins.
            Genes without protein or EST similarity, that are predicted by more
            than two gene prediction programs over most of their length are
            annotated as 'hypothetical' proteins. Genes encoding tRNAs are
            predicted by tRNAscan-SE (Sean Eddy,
            http://genome.wustl.edu/eddy/tRNAscan-SE/). Simple repeats are
            identified by repeatmasker (Arian Smit,
            http://ftp.genome.washington.edu/RM/RepeatMasker.html).
FEATURES             Location/Qualifiers
     source          1. .91924
                     /organism="Arabidopsis thaliana"
                     /cultivar="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="III"
                     /map="mi172"
                     /clone="F22F7"
     misc_feature    1. .14310
                     /note="the annotation for genes within this region can be
                     found in the overlapping bac F18C1 sequence 86291-100600"
     mRNA            complement(join(14688. .14927,15027. .15167,15254. .15373,
                     15468. .15555,15643. .15761,15940. .16131,16211. .16402,
                     16508. .16615,16717. .16833,17184. .>17303))
                     /gene="F22F7.1"
     gene            complement(14688. .>17303)
                     /gene="F22F7.1"
                     /note="identical to GB:AAF22525 from [Arabidopsis
                     thaliana]"
     CDS             complement(join(14850. .14927,15027. .15167,15254. .15373,
                     15468. .15555,15643. .15761,15940. .16131,16211. .16402,
                     16508. .16615,16717. .16833,17184. .17303))
                     /gene="F22F7.1"
                     /codon_start=1
                     /product="26S proteasome AAA-ATPase subunit RPT5a"
                     /protein_id="AAF64530.1"
                     /db_xref="GI:7596759"
                     /translation="MATPMVEDTSSFEEDQLASMSTEDITRATRLLDNEIRILKEDAQ
                     RTNLECDSYKEKIKENQEKIKLNKQLPYLVGNIVEILEMNPEDDAEEDGANIDLDSQR
                     KGKCVVLKTSTRQTIFLPVVGLVDPDSLKPGDLVGVNKDSYLILDTLPSEYDSRVKAM
                     EVDEKPTEDYNDIGGLEKQIQELVEAIVLPMTHKERFEKLGVRPPKGVLLYGPPGTGK
                     TLMARACAAQTNATFLKLAGPQLVQMFIGDGAKLVRDAFQLAKEKAPCIIFIDEIDAI
                     GTKRFDSEVSGDREVQRTMLELLNQLDGFSSDERIKVIAATNRADILDPALMRSGRLD
                     RKIEFPHPTEEARARILQIHSRKMNVHPDVNFEELARSTDDFNGAQLKAVCVEAGMLA
                     LRRDATEVNHEDFNEGIIQVQAKKKASLNYYA"
     repeat_region   complement(15902. .15938)
                     /rpt_family="(TA)n"
     repeat_region   complement(18181. .18757)
                     /note="ATREP6|ATREP6 ATREP6 repeat - a consensus."
     repeat_region   complement(18184. .18375)
                     /note="ATREP9|ATREP9 ATREP9 repeat - a consensus."
     repeat_region   complement(19305. .19353)
                     /rpt_family="POLY_A"
     tRNA            19434. .19505
                     /gene="F22F7.2"
                     /product="tRNA-Gly"
                     /anticodon=(pos:19467. .19469,aa:Gly)
```

```
gene            19434. .19505
                /gene="F22F7.2"
mRNA            complement(join(19554. .19816,19914. .19949,20031. .20105,
                20246. .20323,20425. .20503,20596. .20757,20987. .21051,
                21140. .21229,21832. .21947,22115. .22286))
                /gene="F22F7.3"
gene            complement(19554. .22286)
                /gene="F22F7.3"
                /note="identical to GB:CAA05054 from [Arabidopsis
                thaliana]"
CDS             complement(join(19709. .19816,19914. .19949,20031. .20105,
                20246. .20323,20425. .20503,20596. .20757,20987. .21051,
                21140. .21229,21832. .21947,22115. .22232))
                /gene="F22F7.3"
                /codon_start=1
                /product="alpha subunit of F-actin capping protein"
                /protein_id="AAF64531.1"
                /db_xref="GI:7596760"
                /translation="MADEEDELLETELSYDQKKEIAKWFFLNAPAGEINYVAKDLKAV
                LSDEEVYNEAAMEAFPVYNKTHMICLEMPSGAGDVIVSSYSEINENEYLDPRTAQVAI
                VDHVKQICTKVRPANDEELPSLYIEEYRYALDAEIQRYVSESYPKGMSAVNCVKGKDN
                EGPGSDFELVVIITAMRLSPQNFCNGSWRSVWNIDFQDESQVLDIKGKLQVGAHYFEE
                GNVELDAKKDFQDSTIFQSADDCAIAIANIIRHHETEYLASLEVAYSKLPDNTFKDLR
                RKLPVTRTLFPWQNTLQFSLTREVEKELGLGK"
mRNA            complement(join(21860. .21947,22051. .22286))
                /gene="F22F7.3"
CDS             complement(22107. .22232)
                /gene="F22F7.3"
                /note="unknown protein"
                /codon_start=1
                /protein_id="AAF64545.1"
                /db_xref="GI:7596774"
                /translation="MADEEDELLETELSYDQKKEIAKWFFLNAPAGEINYVAKGT"
mRNA            complement(join(<22769. .23281,23762. .23867,23966. .24084,
                24186. .24383,24484. .24549,24636. .24746,25154. .>25267))
                /gene="F22F7.4"
gene            complement(<22769. .>25267)
                /gene="F22F7.4"
                /note="predicted by genemark.hmm, contains Pfam
                profile:PF01553 Acyltransferase"
CDS             complement(join(22769. .23281,23762. .23867,23966. .24084,
                24186. .24383,24484. .24549,24636. .24746,25154. .25267))
                /gene="F22F7.4"
                /note="hypothetical protein"
                /codon_start=1
                /protein_id="AAF64532.1"
                /db_xref="GI:7596761"
                /translation="MGIHFVDKADLWKSALLFNLKLRDRFRIAVDDHRGRATDLTAEE
                ESALFRMLQTVAVPLIGNACHVFMNGFNRVQVYGLEKLHDALLNRPKNKPLVTVSNHV
                ASVDDPFVIASLLPPKFLLDARNLRWTLCATDRCFKNPVTSAFSRSVKVLPISRGEGI
                YQQGMDIAISKLNNGGWVHIFPEGSRSRDGGKTMGSAKRGIGRLILDADTLPMVVPFV
                HTGMQDIMPVGASVPRIGKTVTVIIGDPIHFNDILSTEGAQHVSRKHLYDAVSSRIGQ
                RLYDLKAQVDRVYIEQQSMMSHNAKTPSDRAAEIFHRVDWDSFGMGAQFSEESSPSSK
                PIGQSDDRIVRSPKRRVSPEGGVSLKIKKLMDSTEMMGFAARGLLMNEYKSRVESANV
                GRPLKAWREYFMNRGL"
mRNA            complement(join(25975. .26514,26764. .26988,27262. .27395))
                /gene="F22F7.5"
```

```
     gene            complement(25975. .27395)
                     /gene="F22F7.5"
                     /note="similar to stress related protein GB:AAD51854 from
                     [Vitis riparia]"
     CDS             complement(join(26041. .26514,26764. .26988,27262. .27303))
                     /gene="F22F7.5"
                     /codon_start=1
                     /product="stress related protein, putative"
                     /protein_id="AAF64533.1"
                     /db_xref="GI:7596762"
                     /translation="MATQTDLAQPKLDMTKEEKERLKYLQFVQAAAVEALLRFALIYA
                     KAKDKSGPLKPGVESVEGAVKTVVGPVYEKYHDVPVEVLKYMDQKVDMSVTELDRRVP
                     PVVKQVSAQAISAAQIAPIVARALASEVRRAGVVETASGMAKSVYSKYEPAAKELYAN
                     YEPKAEQCAVSAWKKLNQLPLFPRLAQVAVPTAAFCSEKYNDTVVKAAEKGYRVTSYM
                     PLVPTERISKIFAEEKAETEPLEFHPLD"
     repeat_region   27792. .27922
                     /rpt_family="(CAAAA)n"
     repeat_region   27796. .27842
                     /rpt_family="POLY_A"
     mRNA            complement(<29103. .29551)
                     /gene="F22F7.6"
     gene            complement(<29103. .29551)
                     /gene="F22F7.6"
     CDS             complement(29103. .29462)
                     /gene="F22F7.6"
                     /note="unknown protein"
                     /codon_start=1
                     /protein_id="AAF64534.1"
                     /db_xref="GI:7596763"
                     /translation="MTNTRAIYAVIAILAIVISAVESTGDFGDSLDFVRAGSSSLFSG
                     CTGSIAECIAEEEEMEFDSDISRRILAQKKYISYGAMRRNSVPCSRRGASYYNCQRGA
                     QANPYSRGCSTITRCRR"
     repeat_region   complement(30238. .30318)
                     /rpt_family="POLY_A"
     repeat_region   31392. .31423
                     /rpt_family="POLY_A"
     mRNA            complement(join(<32605. .32726,32816. .32856,33102. .33246,
                     33331. .33627,34033. .34104,34188. .34439,34681. .34789,
                     34872. .34991,35356. .>35460))
                     /gene="F22F7.7"
     gene            complement(<32605. .>35460)
                     /gene="F22F7.7"
                     /note="predicted by genemark.hmm"

  Query Match             14.2%;  Score 59.8;  DB 8;  Length 91924;
  Best Local Similarity   76.8%;  Pred. No. 1.3e-05;
  Matches   73;  Conservative    0;  Mismatches   22;  Indels    0;  Gaps    0;

  Qy     164 aggttgcctctggctaggtgggatcagcagttcgatcgcctacaactggtcgcggcccaa 223
             ||||||  | |||||| |||| || | | |||| |||||||| |||||||| |  ||
  Db    7608 AGGTTGTTTATGGCTAAGTGGTATTACCGGTTCTATCGCCTATAACTGGTCCCAACCTGC 7549

  Qy     224 tatgaagcctagcgtcaagatcatccacgcaaggt 258
             |||||  | || ||||||||||||||||| ||||
  Db    7548 CATGAAAACCAGTGTCAAGATCATCCACGCCAGGT 7514
```

RESULT 2
ATAC011620/c
LOCUS ATAC011620 100887 bp DNA PLN 24-JAN-2001
DEFINITION Arabidopsis thaliana chromosome III BAC F18C1 genomic sequence, complete sequence.
ACCESSION AC011620
VERSION AC011620.8 GI:12408732
KEYWORDS HTG.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 100887)
AUTHORS Lin,X., Kaul,S., Town,C.D., Benito,M., Creasy,T.H., Haas,B., Wu,D., Ronning,C.M., Koo,H., Fujii,C.Y., Utterback,T.R., Barnstead,M.E., Bowman,C.L., White,O., Nierman,W.C. and Fraser,C.M.
TITLE Arabidopsis thaliana chromosome III BAC F18C1 genomic sequence
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 100887)
AUTHORS Lin,X. and Kaul,S.
TITLE Direct Submission
JOURNAL Submitted (08-OCT-1999) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA, xlin@tigr.org
REFERENCE 3 (bases 1 to 100887)
AUTHORS Lin,X.
TITLE Direct Submission
JOURNAL Submitted (24-JAN-2001) The Institute for Genomic Research, 9712 Medical Center Dr., Rockville, MD 20850, USA
COMMENT On Jan 24, 2001 this sequence version replaced gi:12280779.
Address all correspondence to:at@tigr.org

BAC clone F18C1 is from Arabidopsis chromosome III and is near the molecular marker mi172.
The orientation of the sequence is from SP6 to T7 end of the BAC clone.
Genes were identified by a combination of three methods: Gene prediction programs including GRAIL (available by anonymous ftp from arthur.epm.ornl.gov), Genefinder (Phil Green, University of Washington), Genscan (Chris Burge, http://gnomic.stanford.edu/~chris/GENSCANW.html), and NetPlantGene (http://www.cbs.dtu.dk/netpgene/cbsnetpgene.html), searches of the complete sequence against a peptide database and the Arabidopsis EST database at TIGR (http://www.tigr.org/tdb/at/at.html). Annotated genes are named to indicate the level of evidence for their annotation. Genes with similarity to other proteins are named after the database hits. Genes without significant peptide similarity but with EST similarity are named as 'unknown' proteins. Genes without protein or EST similarity, that are predicted by more than two gene prediction programs over most of their length are annotated as 'hypothetical' proteins. Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, http://genome.wustl.edu/eddy/tRNAscan-SE/). Simple repeats are identified by repeatmasker (Arian Smit, http://ftp.genome.washington.edu/RM/RepeatMasker.html).
FEATURES Location/Qualifiers
source 1. .100887

```
                     /organism="Arabidopsis thaliana"
                     /cultivar="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="III"
                     /map="mi172"
                     /clone="F18C1"
     misc_feature    complement(1. .12001)
                     /note="overlap with BAC clone F10A16 (AC012393:1. .12001)."
     repeat_region   2971. .3040
                     /rpt_family="(GAA)n"
     repeat_region   complement(5045. .5109)
                     /rpt_family="(GAA)n"
     repeat_region   complement(14590. .14654)
                     /rpt_family="POLY_A"
     mRNA            join(<15397. .15585,16169. .16247,16449. .16543,16641.
.16772,
                     17050. .17235,17680. .17843,17919. .17997,18184. .18312,
                     18413. .18540,18736. .>19141)
                     /gene="F18C1.1"
     gene            <15397. .>19141
                     /gene="F18C1.1"
                     /note="similar to GB:AAC27644"
     CDS             join(15397. .15585,16169. .16247,16449. .16543,16641.
.16772,
                     17050. .17235,17680. .17843,17919. .17997,18184. .18312,
                     18413. .18540,18736. .19141)
                     /gene="F18C1.1"
                     /codon_start=1
                     /product="putative importin alpha"
                     /protein_id="AAF26125.1"
                     /db_xref="GI:6714438"
                     /translation="MKGGETMSVRRSGYKAVVDGVGGRRRREDDMVEIRKAKREESLL
                     KKRREALPHSPSADSLDQKLISCIWSDERDLLIEATTQIRTLLCGEMFNVRVEEVIQA
                     GLVPRFVEFLTWDDSPQLQFEAAWALTNIASGTSENTEVVIDHGAVAILVRLLNSPYD
                     VVREQVVWALGNISGDSPRCRDIVLGHAALPSLLLQLNHGAKLSMLVNAAWTLSNLCR
                     GKPQPPFDQVSAALPALAQLIRLDDKELLAYTCWALVYLSDGSNEKIQAVIEANVCAR
                     LIGLSIHRSPSVITPALRTIGNIVTGNDSQTQHIIDLQALPCLVNLLRGSYNKTIRKE
                     ACWTVSNITAGCQSQIQAVFDADICPALVNLLQNSEGDVKKEAAWAICNAIAGGSYKQ
                     IMFLVKQECIKPLCDLLTCSDTQLVMVCLEALKKILKVGEVFSSRHAEGIYQCPQTNV
                     NPHAQLIEEAEGLEKIEGLQSHENNDIYETAVKILETYWMEEEEEEDQEQQDMIYFPV
                     DNFANMPTSSGTLSEMHCGP"
     repeat_region   19025. .19059
                     /rpt_family="(GAA)n"
     mRNA            complement(join(19632. .20011,20101. .20181,20264. .20329,
                     20424. .20501,20581. .20652,20874. .20963,21057. .21333,
                     21648. .>21871))
                     /gene="F18C1.2"
     gene            complement(19632. .>21871)
                     /gene="F18C1.2"
                     /note="similar to putative syntaxin protein GB:CAB52175
                     from [Arabidopsis thaliana]"
     CDS             complement(join(19904. .20011,20101. .20181,20264. .20329,
                     20424. .20501,20581. .20652,20874. .20963,21057. .21333,
                     21648. .21871))
                     /gene="F18C1.2"
                     /codon_start=1
                     /product="putative syntaxin protein, AtSNAP33"
```

```
                     /protein_id="AAF26126.1"
                     /db_xref="GI:6714439"
                     /translation="MATRNRTLLFRKYRNSLRSVRAPMGSSSSSTLTEHNSLTGAKSG
                     LGPVIEMASTSLLNPNRSYAPVSTEDPGNSSRGTITVGLPPDWVDVSEEISVYIQRAR
                     TKMAELGKAHAKALMPSFGDGKEDQHQIETLTQEVTFLLKKSEKQLQRLSAAGPSEDS
                     NVRKNVQRSLATDLQNLSMELRKKQSTYLKRLRLQKEDGADLEMNLNGSRYKAEDDDF
                     DDMVFSEHQMSKIKKSEEISIEREKEIQQVVESVSELAQIMKDLSALVIDQGTIVDRI
                     DYNIQNVASTVDDGLKQLQKAERTQRQGGMVMCASVLVILCFIMLVLLILKEILL"
     repeat_region   22133. .22223
                     /rpt_family="(TAAA)n"
     repeat_region   22179. .22226
                     /rpt_family="POLY_A"
     mRNA            join(<22810. .22885,23454. .23596,23884. .23922,24150.
.24328,
                     24593. .>24701)
                     /gene="F18C1.3"
     gene            <22810. .>24701
                     /gene="F18C1.3"
     CDS             join(22810. .22885,23454. .23596,23884. .23922,24150.
.24328,
                     24593. .24701)
                     /gene="F18C1.3"
                     /note="unknown protein"
                     /codon_start=1
                     /protein_id="AAF26127.1"
                     /db_xref="GI:6714440"
                     /translation="MDSDSWSDRLASATRRYQLAFPSRSDTFLGFEEIDGEEEFREEF
                     ACPFCSDYFDIVSLCCHIDEDHPMEAKNGVCPVCAVRVGVDMSLFGGSSCIVSSSSSS
                     NVAADPLLSSFISPIADGFFTTESCISAETGPVKKTTIQCLPEQNAKKTSLSAEDHKQ
                     KLKRSEFVRELLSSTILDDSL"
     repeat_region   complement(27249. .27313)
                     /rpt_family="(TAAAA)n"
     mRNA            join(28290. .28335,28744. .29013,29330. .29424,29536.
.29754,
                     29843. .30597)
                     /gene="F18C1.4"
     gene            28290. .30597
                     /gene="F18C1.4"
                     /note="similar to GB:CAA74049"
     CDS             join(28809. .29013,29330. .29424,29536. .29754,29843.
.30211)
                     /gene="F18C1.4"
                     /codon_start=1
                     /product="putative transcription factor"
                     /protein_id="AAF26128.1"
                     /db_xref="GI:6714441"
                     /translation="MAMQTVREGLFSAPQTSWWTAFGSQPLAPESLAGDSDSFAGVKV
                     GSVGETGQRVDKQSNSATHLAFSLGDVKSPRLVPKPHGATFSMQSPCLELGFSQPPIY
                     TKYPYGEQQYYGVVSAYGSQSRVMLPLNMETEDSTIYVNSKQYHGIIRRRQSRAKAAA
                     VLDQKKLSSRCRKPYMHHSRHLHALRRPRGSGGRFLNTKSQNLENSGTNAKKGDGSMQ
                     IQSQPKPQQSNSQNSEVVHPENGTMNLSNGLNVSGSEVTSMNYFLSSPVHSLGGMVMP
                     SKWIAAAAAMDNGCCNFKT"
     mRNA            join(36087. .36254,36777. .37418,37507. .37875,37965.
.38069,
                     38171. .38353,38529. .38670,38758. .38939,39026. .39111,
                     39212. .39260,39735. .39801,39900. .40385,40434. .40540,
                     41173. .41240,41336. .41552,41793. .42008,42111. .42266,
```

```
                     42340. .42393,42479. .42687,42782. .42845,43070. .43389,
                     43742. .43925,44024. .46030,46087. .46203,46311. .46581)
                     /gene="F18C1.5"
     gene            36087. .46581
                     /gene="F18C1.5"
     CDS             join(36093. .36254,36777. .37418,37507. .37875,37965.
.38069,
                     38171. .38353,38529. .38670,38758. .38939,39026. .39111,
                     39212. .39260,39735. .39801,39900. .40385,40434. .40540,
                     41173. .41240,41336. .41552,41793. .42008,42111. .42266,
                     42340. .42393,42479. .42687,42782. .42845,43070. .43389,
                     43742. .43925,44024. .46030,46087. .46203,46311. .46331)
                     /gene="F18C1.5"
                     /note="unknown protein"
                     /codon_start=1
                     /protein_id="AAF26129.1"
                     /db_xref="GI:6714442"
                     /translation="MVRYLNVDNIFVRATSPPSFALEVFVRCEGESKFKRLCNPFLYT
                     PSAPYPLEVEAVVTNHLVVRGSYRSLSLIVYGNIVKDLGQYNIILEGRSVTDIVSSTE
                     GNLEDLPLVLHSVNRTIEECLSSLDIVSLPLAAVDLPVEVKRLLQLLLKIFDKLATND
                     VVNKFVDTVVSGVSSYVTDNVDFFLKNKNCSAVTSSLDSGLFHDIVDRVKEDILDLNE
                     IQESDVALGLFSFLESETYLATSQQLVVMLSPYIQFERDSLCTVLPKLSKGKATLLGL
                     SLAFLLCSGREGCLQFVNSGGMDQLVYLFGHDGQNSTTITLLLLGVVEQATRHSVGCE
                     GFLGWWPREDGSIPSGKSEGYCLLLKLLMQKPCHEIASLAIYILRRLRIYEVISRYEF
                     AVLSALEGLSNSHGAATHNLNMLSDAKSQLQKLQNLMKSLGSVEDPSPSAYAERSLVS
                     DHSEGWLSYKATSKLTSSWTCPFYSSGIDSHILALLKERGFLPLSAALLSMPELHSKV
                     GDIMDVFTDIAMFIGNIILSFMFSRTGLSFLLHHPELTATIIQSLKGSVDLNKEECVP
                     LHYASILISKGFTCSLLEIGINLEMHLRVVSAVDRLLKSIQQTEEFLWILWELRDVSR
                     SDCGREALLTLGVFPEALAVLIEALHSAKDMEPAVENSGISPLNLAICHSAAEIFEVI
                     VSDSTASCLHAWIEHAPVLHKALHTLSPGGSNRKDAPSRLLKWIDAGVVYHKHGVGGL
                     LRYAAVLASGGDAQLSSSSILALDLTPAENGAGESTNVSEMNVLDNLGKVIFEKSFEG
                     VNLSDSSISQLTTALRILALISDNSVYAIVFLFKCSEMLFFVQTVAAALYDEGAVTVV
                     YAILGTKEQYRNTKLMKALLRLHREVSPKLAACAADLSSHYPDSALGFGAVCHLIVSA
                     LVCWPVYGWIPGLFHTLLSGVQTSSVPALGPKETCSFLCILSDILPEEGVWFWKSGMP
                     LLSGLRKLAVGTLMGPQKEKQINWYLEPGPLEKLINHLTPNLDKIAKIIQHHAVSALV
                     VIQDMLRVFIVRIACQRVEHASILLRPIFSSIRDGILDQSSTRDTEAYMVYRYLNFLA
                     SLLEHPHAKGLLLEEGIVQLLVEVLERCYDATYPSENRVLEYGIVSASSVIQWCIPAF
                     RSISLLCDSQVPLLCFQKKELLASLSAKDCALIFPFVLKFCQVLPVGNELLSCLGAFK
                     DLSSCGEGQDGLVSLLFHLFSGTEESVSERWCDTNSLSLDQLDMKKNPPFLSCWIKLL
```

```
  Query Match             14.2%;  Score 59.8;  DB 8;  Length 100887;
  Best Local Similarity   76.8%;  Pred. No. 1.3e-05;
  Matches   73;  Conservative    0;  Mismatches   22;  Indels    0;  Gaps    0;

Qy    164 aggttgcctctggctaggtgggatcagcagttcgatcgcctacaactggtcgcggcccaa 223
          ||||||  | |||||| |||| || | | |||| |||||||| |||||||| |  ||
Db  93898 AGGTTGTTTATGGCTAAGTGGTATTACCGGTTCTATCGCCTATAACTGGTCCCAACCTGC 93839

Qy    224 tatgaagcctagcgtcaagatcatccacgcaaggt 258
           |||||  | || |||||||||||||||| ||||
Db  93838 CATGAAAACCAGTGTCAAGATCATCCACGCCAGGT 93804
```

```
RESULT   3
G69921
LOCUS       G69921        255 bp    DNA              STS       08-JUN-2001
DEFINITION  695251131FN194 maize leaf DNA Zea mays STS genomic, sequence tagged
```

site.
ACCESSION G69921
VERSION G69921.1 GI:14331606
KEYWORDS STS.
SOURCE Zea mays.
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 255)
AUTHORS Yang,Y.J., Guo,L., Ashlock,D.A., Wen,T.J. and Schnable,P.S.
TITLE 3' UTR sequences of maize genes
JOURNAL Unpublished
COMMENT
Contact: Schnable, P.S.
Schnable laboratory
Iowa State University
G405 Agronomy Hall, Ames, IA 50011, USA
Tel: 515-294-0975
Fax: 515-294-2299
Email: schnable@iastate.edu
Primer A: CGAGGGAGTGGGTGGTC
Primer B: TGGTACCATCTGCATACACAAC
PCR Profile:
Denaturation: 94 degrees C for 30 seconds
Annealing: 60 degrees C for 45 seconds
Polymerization: 72 degrees C for 90 seconds
PCR cycles: 31
Thermal cycler: Perkin Elmer TC
Protocol:
Template: 10-20 ng
Primer: each 0.5 uM
dNTPs: each 200 uM
Taq Polymerase: 0.05 units/ul
Total vol: 20 ul

Buffer:
MgCl2: 2 mM
KCl: 50 mM
Tris-HCl: 20 mM
pH: 8.4.
FEATURES Location/Qualifiers
source 1. .255
/organism="Zea mays"
/strain="DE811"
/db_xref="taxon:4577"
/clone_lib="maize leaf DNA"
/note="PCR products amplified from genomic DNA"
STS <1. .>255
BASE COUNT 46 a 82 c 65 g 62 t
ORIGIN

Query Match 13.3%; Score 56; DB 11; Length 255;
Best Local Similarity 71.2%; Pred. No. 0.00023;
Matches 74; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

```
Qy     160 ccgtaggttgcctctggctaggtgggatcagcagttcgatcgcctacaactggtcgcggc 219
           | | |||    || |||||  | || |||  | | || ||||||||||||||||  |||
Db      66 CTGCAGGGACGCTGTGGCTCAGCGGCATCGTCGGCTCCATCGCCTACAACTGGTCCAGGC 125

Qy     220 ccaatatgaagcctagcgtcaagatcatccacgcaaggttgcat 263
           |    |||||  | | ||||||||||||||||||| ||||  | |
Db     126 CGGGCATGAAAACCAACGTCAAGATCATCCACGCCAGGTCCCCT 169
```

```
RESULT   4
G69610
LOCUS       G69610         387 bp    DNA              STS       08-JUN-2001
DEFINITION  695251131FB113 maize leaf DNA Zea mays STS genomic, sequence tagged
            site.
ACCESSION   G69610
VERSION     G69610.1  GI:14331295
KEYWORDS    STS.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 387)
  AUTHORS   Yang,Y.J., Guo,L., Ashlock,D.A., Wen,T.J. and Schnable,P.S.
  TITLE     3' UTR sequences of maize genes
  JOURNAL   Unpublished
COMMENT
            Contact: Schnable, P.S.
            Schnable laboratory
            Iowa State University
            G405 Agronomy Hall, Ames, IA 50011, USA
            Tel: 515-294-0975
            Fax: 515-294-2299
            Email: schnable@iastate.edu
            Primer A: CGAGGGAGTGGGTGGTC
            Primer B: TGGTACCATCTGCATACACAAC
            PCR Profile:
                 Denaturation:   94 degrees C for 30 seconds
                 Annealing:      60 degrees C for 45 seconds
                 Polymerization: 72 degrees C for 90 seconds
                 PCR cycles:     31
                 Thermal cycler: Perkin Elmer TC
            Protocol:
                  Template: 10-20 ng
                  Primer: each 0.5 uM
                  dNTPs: each 200 uM
                  Taq Polymerase: 0.05 units/ul
                  Total vol: 20 ul

            Buffer:
                  MgCl2: 2 mM
                  KCl: 50 mM
                  Tris-HCl: 20 mM
                  pH: 8.4.
FEATURES             Location/Qualifiers
     source          1. .387
                     /organism="Zea mays"
```

```
                     /strain="DE811"
                     /db_xref="taxon:4577"
                     /clone_lib="maize leaf DNA"
                     /note="PCR products amplified from genomic DNA"
     STS             <1. .>387
BASE COUNT       79 a    106 c     91 g    111 t
ORIGIN


  Query Match             13.3%;  Score 56;  DB 11;  Length 387;
  Best Local Similarity   71.2%;  Pred. No. 0.00022;
  Matches   74;  Conservative    0;  Mismatches   30;  Indels    0;  Gaps    0;

Qy     160 ccgtaggttgcctctggctaggtgggatcagcagttcgatcgcctacaactggtcgcggc 219
           | | |||    || |||||  | || |||  | | || ||||||||||||||| |  |||
Db      87 CTGCAGGGACGCTGTGGCTCAGCGGCATCGTCGGCTCCATCGCCTACAACTGGGCCAGGC 146

Qy     220 ccaatatgaagcctagcgtcaagatcatccacgcaaggttgcat 263
           |    |||||| | ||||| |||||||||||||| ||||  | |
Db     147 CGGGCATGAAGACCAGCGTTAAGATCATCCACGCCAGGTCCCCT 190


RESULT   5
G70259
LOCUS       G70259         368 bp    DNA              STS       08-JUN-2001
DEFINITION  695251131FB73-tj maize leaf DNA Zea mays STS genomic, sequence
            tagged site.
ACCESSION   G70259
VERSION     G70259.1  GI:14331944
KEYWORDS    STS.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 368)
  AUTHORS   Yang,Y.J., Guo,L., Ashlock,D.A., Wen,T.J. and Schnable,P.S.
  TITLE     3' UTR sequences of maize genes
  JOURNAL   Unpublished
COMMENT
            Contact: Schnable, P.S.
            Schnable laboratory
            Iowa State University
            G405 Agronomy Hall, Ames, IA 50011, USA
            Tel: 515-294-0975
            Fax: 515-294-2299
            Email: schnable@iastate.edu
            Primer A: CGAGGGAGTGGGTGGTC
            Primer B: TGGTACCATCTGCATACACAAC
            PCR Profile:
                 Denaturation:   94 degrees C for 30 seconds
                 Annealing:      60 degrees C for 45 seconds
                 Polymerization: 72 degrees C for 90 seconds
                 PCR cycles:     31
                 Thermal cycler: Perkin Elmer TC
            Protocol:
                   Template: 10-20 ng
```

```
                    Primer: each 0.5 uM
                    dNTPs: each 200 uM
                    Taq Polymerase: 0.05 units/ul
                    Total vol: 20 ul

               Buffer:
                     MgCl2: 2 mM
                     KCl: 50 mM
                     Tris-HCl: 20 mM
                     pH: 8.4.
FEATURES             Location/Qualifiers
     source          1. .368
                     /organism="Zea mays"
                     /strain="DE811"
                     /db_xref="taxon:4577"
                     /clone_lib="maize leaf DNA"
                     /note="PCR products amplified from genomic DNA"
     STS             <1. .>368
BASE COUNT       80 a     96 c     86 g    106 t
ORIGIN


  Query Match             12.9%;  Score 54.4;  DB 11;  Length 368;
  Best Local Similarity   70.2%;  Pred. No. 0.0006;
  Matches   73;  Conservative    0;  Mismatches   31;  Indels    0;  Gaps    0;

Qy     160 ccgtaggttgcctctggctaggtgggatcagcagttcgatcgcctacaactggtcgcggc 219
           | | |||    || |||||  | || |||  | | || ||||||||||||||||||  |||
Db      68 CTGCAGGGACGCTGTGGCTCAGCGGCATCGTCGGCTCCATCGCCTACAACTGGTCCAGGC 127

Qy     220 ccaatatgaagcctagcgtcaagatcatccacgcaaggttgcat 263
           |    |||||  | | ||| |||||||||||||| ||||  | |
Db     128 CGGGCATGAAAACCATCGTTAAGATCATCCACGCCAGGTCCCCT 171


RESULT   6
AF180335/c
LOCUS       AF180335    27588 bp    DNA             PLN       28-SEP-1999
DEFINITION  Glycine max malate dehydrogenase (Mdh1) gene, complete cds; nuclear
            gene for mitochondrial product.
ACCESSION   AF180335
VERSION     AF180335.1  GI:5929963
KEYWORDS    .
SOURCE      soybean.
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Glycine.
REFERENCE   1  (bases 1 to 27588)
  AUTHORS   Imsande,J., Pittig,J., Palmer,R.G. and Gietl,C.
  TITLE     Independent spontaneously induced mitochondrial malate
            dehydrogenase null mutants in soybean are the result of a deletion
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 27588)
  AUTHORS   Imsande,J., Pittig,J., Palmer,R.G. and Gietl,C.
  TITLE     Direct Submission
```

```
  JOURNAL   Submitted (24-AUG-1999) Plant Genetics, Agronomy, Iowa State
            University, Ames, IA 50011, USA
FEATURES             Location/Qualifiers
     source          1. .27588
                     /organism="Glycine max"
                     /cultivar="Williams"
                     /db_xref="taxon:3847"
     repeat_region   981. .1135
                     /rpt_type=dispersed
     misc_feature    4113. .4345
                     /note="similar to AMP binding proteins deposited in
                     GenBank Accession Number CAB10186 and encoded by GenBank
                     Accession Number Z72151"
     repeat_region   5039. .5235
                     /rpt_type=dispersed
     repeat_region   5865. .6065
                     /rpt_type=dispersed
     mRNA            join(<7254. .7508,8542. .8715,9167. .9335,9616. .9677,
                     9783. .9904,10062. .10179,10567. .>10704)
                     /gene="Mdh1"
                     /product="malate dehydrogenase"
     gene            <7254. .>10704
                     /gene="Mdh1"
     CDS             join(7254. .7508,8542. .8715,9167. .9335,9616. .9677,
                     9783. .9904,10062. .10179,10567. .10704)
                     /gene="Mdh1"
                     /codon_start=1
                     /product="malate dehydrogenase"
                     /protein_id="AAD56659.1"
                     /db_xref="GI:5929964"
                     /translation="MMKPSMLRSLHSAATRGASHLFRRGYASEPVPERKVAVLGAAGG
                     IGQPLSLLMKLNPLVSSLSLYDIAGTPGVAADISHINTRSEVVGYQGDEELGKALEGA
                     DVVIIPAGVPRKPGMTRDDLFNINAGIVKTLCTAIAKYCPHALVNMISNPVNSTVPIA
                     AEVFKKAGTYDEKRLFGVTTLDVVRAKTFYAGKANVPVAGVNVPVVGGHAGITILPLF
                     SQATPKANLDDDVIKALTKRTQDGGTEVVEAKAGKGSATLSMAYAGALFADACLKGLN
                     GVPDVVECSFVQSTVTELPFFASKVRLGTVGVEEVLGLGHLSDFEQQGLESLKPELKS
                     SIEKGIKFANQ"
     repeat_region   11875. .12241
                     /note="similar to a barley retrotransposon"
                     /rpt_type=dispersed
     repeat_region   15742. .16109
                     /note="similar to a barley retrotransposon"
                     /rpt_type=dispersed
     misc_feature    complement(join(16851. .17206,17865. .18164))
                     /note="similar to antisense of soybean catalase encoded by
                     GenBank Accession Number Z12021"
     repeat_region   18120. .18135
                     /rpt_type=tandem
                     /rpt_unit=1812. .18127
     repeat_region   18141. .18156
                     /rpt_type=tandem
                     /rpt_unit=18141. .18148
     misc_feature    complement(26350. .27063)
                     /gene="gag-pol"
                     /note="gag-pol protein with internal stop codon;
                     transposition of retrotransposon"
     repeat_region   complement(<26350. .>27063)
```

```
                     /note="retrotransposon"
                     /rpt_type=dispersed
     gene            complement(26350. .27063)
                     /gene="gag-pol"
BASE COUNT     8635 a    5055 c    5183 g    8665 t      50 others
ORIGIN


  Query Match             12.4%;  Score 52;  DB 8;  Length 27588;
  Best Local Similarity   70.0%;  Pred. No. 0.0017;
  Matches   70;  Conservative    0;  Mismatches   30;  Indels    0;  Gaps    0;

Qy     164 aggttgcctctggctaggtgggatcagcagttcgatcgcctacaactggtcgcggcccaa 223
           ||| || || |||||  |||| ||     |||| ||||| ||||| ||||| || |||||
Db    6194 AGGGTGTCTGTGGCTCAGTGGTATTTCGCGTTCAATCGCTTACAATTGGTCTCGACCCAA 6135

Qy     224 tatgaagcctagcgtcaagatcatccacgcaaggttgcat 263
           |||||     | || | |||||| |||||||| |   ||
Db    6134 CATGAAAATCGGTGTTAGGATCATTCACGCAAGTTCCTAT 6095


RESULT   7
AC069556
LOCUS       AC069556    93672 bp    DNA             PLN       27-JUL-2000
DEFINITION  Genomic Sequence For Arabidopsis thaliana Clone T1G16 From
            Chromosome V, complete sequence.
ACCESSION   AC069556
VERSION     AC069556.4  GI:9502396
KEYWORDS    HTG.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 93672)
  AUTHORS   Rodriguez,M.A., Nascimento,L.U., de la Bastide,M., Preston,R.R.,
            Huang,E.N., See,L.H., Spiegel,L.A., Baker,J.P., Vil,M.D.,
            Shah,R.S., Bahret,A., King,L., Kirchoff,K.A., Miller,B.,
            Shekher,M., Toth,K., O'Shaughnessy,A., Dedhia,N.N. and
            McCombie,W.R.
  TITLE     Genomic Sequence For Arabidopsis thaliana Clone T1G16 From
            Chromosome V, Complete Sequence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 93672)
  AUTHORS   McCombie,W.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-JUN-2000) Lita Annenberg Hazen Genome Sequencing
            Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring
            Harbor, NY 11724, USA
REFERENCE   3  (bases 1 to 93672)
  AUTHORS   McCombie,W.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JUL-2000) Lita Annenberg Hazen Genome Sequencing
            Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring
            Harbor, NY 11724, USA
COMMENT     On Jul 27, 2000 this sequence version replaced gi:8698718.
            This sequence was finished as follows unless otherwise noted: all
```

regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest.
Bases 1-18896 overlap with F15A18 (Accession # AC007478). The overlap is from 89035-107931 on F15A18. Bases 90723-93672 overlap with F14I23 (Accession # AC007399). The overlap is from 1-2949 on F14I23.

```
FEATURES             Location/Qualifiers
     source          1. .93672
                     /organism="Arabidopsis thaliana"
                     /db_xref="taxon:3702"
                     /chromosome="V"
                     /clone="T1G16"
                     /clone_lib="TAMU"
     misc_feature    25520. .25590
                     /note="This region is covered by one PCR product of high
                     quality data."
BASE COUNT    30853 a  15991 c  16091 g  30737 t
ORIGIN


  Query Match             12.2%;  Score 51.2;  DB 8;  Length 93672;
  Best Local Similarity   70.8%;  Pred. No. 0.0024;
  Matches   68;  Conservative    0;  Mismatches   28;  Indels    0;  Gaps    0;

Qy    163 taggttgcctctggctaggtgggatcagcagttcgatcgcctacaactggtcgcggccca 222
          |||||||| | |||||| |||| |||    |||| || || || || |||||    ||
Db  51044 TAGGTTGCTTATGGCTAAGTGGTATCTCTGGTTCAATTGCTTATAATTGGTCTAAACCTG 51103

Qy    223 atatgaagcctagcgtcaagatcatccacgcaaggt 258
            |||||  | || ||||  ||||||||||| ||||
Db  51104 CCATGAAAACCAGTGTCAGAATCATCCACGCTAGGT 51139


RESULT   8
I66494/c
LOCUS       I66494       7218 bp    DNA             PAT       28-DEC-1997
DEFINITION  Sequence 14 from patent US 5670367.
ACCESSION   I66494
VERSION     I66494.1  GI:2724471
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 7218)
  AUTHORS   Dorner,F., Scheiflinger,F. and Falkner,F.Gunter.
  TITLE     Recombinant fowlpox virus
  JOURNAL   Patent: US 5670367-A 14 23-SEP-1997;
FEATURES             Location/Qualifiers
     source          1. .7218
                     /organism="unknown"
BASE COUNT     1944 a   1491 c   1486 g   1929 t    368 others
ORIGIN
```

```
  Query Match             11.8%;  Score 49.6;  DB 6;  Length 7218;
  Best Local Similarity   6.4%;  Pred. No. 0.0083;
  Matches   10;  Conservative  106;  Mismatches   40;  Indels    0;  Gaps    0;

Qy      1 gaaaaaaataactcggaaaagaaggagacgccgaaaattcgaaaggggaggggaaagcaa 60
          :::::::: ::   :::::::::::::: :  :::::   ::::::::::::::::: ::
Db   1210 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1151

Qy     61 agctgatggcggaggcccaggggaaagcaaagcaaatggcggaggccccgagcaagatcg 120
          ::  :: :: :::::   ::::::::: :::: ::: :: :::::    ::: ::::  :
Db   1150 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1091

Qy    121 aatccatgaggaagtgggtcgtcgagcacaagctcc 156
          ::   : ::::::: :::  :  :| | ||||||||
Db   1090 RRRRRRRRRRRRRRRRRRRRRRRRATCGCAAGCTCC 1055


RESULT   9
CBRG45O11/c
LOCUS       CBRG45O11   26703 bp    DNA             INV       04-NOV-2000
DEFINITION  Caenorhabditis briggsae cosmid G45O11, complete sequence.
ACCESSION   AC084652
VERSION     AC084652.1  GI:11095098
KEYWORDS    HTG.
SOURCE      Caenorhabditis briggsae.
  ORGANISM  Caenorhabditis briggsae
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
            Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 26703)
  AUTHORS   Washington University Genome Sequencing Center.
  TITLE     The C. briggsae Genome Sequencing Project
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 26703)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-NOV-2000) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     Submitted by:
                    Genome Sequencing Center
                    Department of Genetics, Washington University,
                    St. Louis, MO 63110, USA
                    e-mail: jspieth@watson.wustl.edu


            NOTICE:  This sequence may not be the entire insert of this clone.
            It may be shorter because we only sequence overlapping sections
            once, or longer because we provide a small overlap between
            neighboring submissions.
FEATURES              Location/Qualifiers
     source           1. .26703
                      /organism="Caenorhabditis briggsae"
                      /strain="GujArat G16"
                      /db_xref="taxon:6238"
                      /clone="G45O11"
BASE COUNT     5831 a    6601 c    4766 g    9505 t
ORIGIN
```

```
  Query Match             10.8%;  Score 45.6;  DB 3;  Length 26703;
  Best Local Similarity   56.8%;  Pred. No. 0.083;
  Matches   84;  Conservative    0;  Mismatches   64;  Indels    0;  Gaps    0;

Qy      18 aaagaaggagacgccgaaaattcgaaaggggaggggaaagcaaagctgatggcggaggcc 77
           ||| | | ||  | ||||||  | ||   ||| |   | || |||   | |  ||||||
Db    7364 AAAAAGGAAGCTGACGAAAAGGCCAAGAAGGAAGCCGAGGCTAAGACTAAGAAGGAGGCT 7305

Qy      78 caggggaaagcaaagcaaatggcggaggccccgagcaagatcgaatccatgaggaagtgg 137
            | |  || || | | |   ||| ||||||  ||   ||  |||  ||| || | ||
Db    7304 GATGATAAGGCCAGGAAGGAGGCCGAGGCCAAGAAGGAGGCCGAGGCCAAGAAGGAGGCT 7245

Qy     138 gtcgtcgagcacaagctccgagccgtag 165
           | || | ||  ||||      || ||||
Db    7244 GACGACAAGGCCAAGAAGGAGGCTGTAG 7217


RESULT  10
AC084066
LOCUS       AC084066   235411 bp    DNA             HTG       12-OCT-2000
DEFINITION  Mus musculus clone RP23-321D1, *** SEQUENCING IN PROGRESS ***, 29
            unordered pieces.
ACCESSION   AC084066
VERSION     AC084066.1  GI:10799415
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 235411)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Mouse
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 235411)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-OCT-2000) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     -------Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -------
            Project Information
            Center Project Name: 2351294
            Center clone name: RPCI-23_321D1
            -------
            Summary Statistics
            Consensus quality: 214207 bases at least Q40
            Consensus quality: 223053 bases at least Q30
            Consensus quality: 225208 bases at least Q20
            Estimated insert size: 200000; pulse field gel estimation
            Estimated insert size: 232611; sum-of-contigs estimation
            Quality coverage:  11.56 in Q20 bases; pulse field gel estimation
            Quality coverage:  9.94 in Q20 bases; sum-of-contigs estimation.
```

```
* NOTE: This is a 'working draft' sequence. It currently
* consists of 29 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*        1     1210: contig of 1210 bp in length
*     1211     1310: gap of unknown length
*     1311     3240: contig of 1930 bp in length
*     3241     3340: gap of unknown length
*     3341     4504: contig of 1164 bp in length
*     4505     4604: gap of unknown length
*     4605     6172: contig of 1568 bp in length
*     6173     6272: gap of unknown length
*     6273     7354: contig of 1082 bp in length
*     7355     7454: gap of unknown length
*     7455     8625: contig of 1171 bp in length
*     8626     8725: gap of unknown length
*     8726    10114: contig of 1389 bp in length
*    10115    10214: gap of unknown length
*    10215    12091: contig of 1877 bp in length
*    12092    12191: gap of unknown length
*    12192    14113: contig of 1922 bp in length
*    14114    14213: gap of unknown length
*    14214    15286: contig of 1073 bp in length
*    15287    15386: gap of unknown length
*    15387    17839: contig of 2453 bp in length
*    17840    17939: gap of unknown length
*    17940    20266: contig of 2327 bp in length
*    20267    20366: gap of unknown length
*    20367    22001: contig of 1635 bp in length
*    22002    22101: gap of unknown length
*    22102    24307: contig of 2206 bp in length
*    24308    24407: gap of unknown length
*    24408    26405: contig of 1998 bp in length
*    26406    26505: gap of unknown length
*    26506    29691: contig of 3186 bp in length
*    29692    29791: gap of unknown length
*    29792    34705: contig of 4914 bp in length
*    34706    34805: gap of unknown length
*    34806    39749: contig of 4944 bp in length
*    39750    39849: gap of unknown length
*    39850    45296: contig of 5447 bp in length
*    45297    45396: gap of unknown length
*    45397    51476: contig of 6080 bp in length
*    51477    51576: gap of unknown length
*    51577    59008: contig of 7432 bp in length
*    59009    59108: gap of unknown length
*    59109    66218: contig of 7110 bp in length
*    66219    66318: gap of unknown length
*    66319    76778: contig of 10460 bp in length
*    76779    76878: gap of unknown length
*    76879    98098: contig of 21220 bp in length
*    98099    98198: gap of unknown length
*    98199   113987: contig of 15789 bp in length
```

```
            *   113988   114087: gap of unknown length
            *   114088   132115: contig of 18028 bp in length
            *   132116   132215: gap of unknown length
            *   132216   160184: contig of 27969 bp in length
            *   160185   160284: gap of unknown length
            *   160285   186698: contig of 26414 bp in length
            *   186699   186798: gap of unknown length
            *   186799   235411: contig of 48613 bp in length.
FEATURES             Location/Qualifiers
     source          1. .235411
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="RP23-321D1"
                     /clone_lib="RPCI mouse BAC library 23"
BASE COUNT    62132 a  54445 c  54225 g  61806 t   2803 others
ORIGIN


  Query Match             10.3%;  Score 43.4;  DB 2;  Length 235411;
  Best Local Similarity   55.7%;  Pred. No. 0.26;
  Matches   83;  Conservative    0;  Mismatches   66;  Indels    0;  Gaps    0;

Qy      2 aaaaaaataactcggaaaagaaggagacgccgaaaattcgaaaggggaggggaaagcaaa 61
          | || || ||   | |||||||| ||| |  ||| |   | | | ||||| | | |  ||
Db 180576 AGAAGAAGAAGAAGAAAAAGAAGAAGAAGAAGAAGAAGAGGAGGAGGAGGAGGAGGTGAA 180635

Qy     62 gctgatggcggaggcccaggggaaagcaaagcaaatggcggaggccccgagcaagatcga 121
          |  || || | |||   ||| ||||   ||| | | |  | ||     ||  ||||   |
Db 180636 GGAGAAGGAGAAGGAGAAGGAGAAAAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAA 180695

Qy    122 atccatgaggaagtgggtcgtcgagcaca 150
            | || |||| | |     ||| | |
Db 180696 GAAGAAGAAGAAGAGAGAGAGAGAGAAGA 180724


RESULT  11
AC067840
LOCUS       AC067840   187278 bp    DNA             HTG       28-MAY-2000
DEFINITION  Homo sapiens chromosome 14 clone RP11-633N4 map 14, WORKING DRAFT
            SEQUENCE, 13 unordered pieces.
ACCESSION   AC067840
VERSION     AC067840.2  GI:8099886
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 187278)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 14, clone RP11-633N4
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 187278)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
```

Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L., Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and Zody,M.

TITLE Direct Submission

JOURNAL Submitted (27-APR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

COMMENT On May 28, 2000 this sequence version replaced gi:7651891.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
-------------- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
-------------- Project Information
Center project name: L9398
Center clone name: 633_N_4
-------------- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 180520 bases at least Q40
Consensus quality: 183792 bases at least Q30
Consensus quality: 185077 bases at least Q20
Insert size: 187000; agarose-fp
Insert size: 186078; sum-of-contigs
Quality coverage: 5.0 in Q20 bases; agarose-fp
Quality coverage: 5.0 in Q20 bases; sum-of-contigs
--------------.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 13 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 2734: contig of 2734 bp in length
* 2735 2834: gap of 100 bp
* 2835 7097: contig of 4263 bp in length
* 7098 7197: gap of 100 bp
* 7198 10127: contig of 2930 bp in length
* 10128 10227: gap of 100 bp

```
            *    10228    15117: contig of 4890 bp in length
            *    15118 15217: gap of      100 bp
            *    15218    24716: contig of 9499 bp in length
            *    24717 24816: gap of      100 bp
            *    24817    32540: contig of 7724 bp in length
            *    32541 32640: gap of      100 bp
            *    32641    47120: contig of 14480 bp in length
            *    47121 47220: gap of      100 bp
            *    47221    64489: contig of 17269 bp in length
            *    64490 64589: gap of      100 bp
            *    64590    82156: contig of 17567 bp in length
            *    82157 82256: gap of      100 bp
            *    82257   104149: contig of 21893 bp in length
            *   104150 104249: gap of      100 bp
            *   104250   126861: contig of 22612 bp in length
            *   126862 126961: gap of      100 bp
            *   126962   152105: contig of 25144 bp in length
            *   152106 152205: gap of      100 bp
            *   152206   187278: contig of 35073 bp in length.
FEATURES             Location/Qualifiers
     source          1. .187278
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="14"
                     /map="14"
                     /clone="RP11-633N4"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .2734
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:left"
     misc_feature    2835. .7097
                     /note="assembly_fragment"
     misc_feature    7198. .10127
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    10228. .15117
                     /note="assembly_fragment"
     misc_feature    15218. .24716
                     /note="assembly_fragment"
     misc_feature    24817. .32540
                     /note="assembly_fragment"
     misc_feature    32641. .47120
                     /note="assembly_fragment"
     misc_feature    47221. .64489
                     /note="assembly_fragment"
     misc_feature    64590. .82156
                     /note="assembly_fragment"
     misc_feature    82257. .104149
                     /note="assembly_fragment"
     misc_feature    104250. .126861
                     /note="assembly_fragment"
     misc_feature    126962. .152105
                     /note="assembly_fragment"
     misc_feature    152206. .187278
                     /note="assembly_fragment"
```

BASE COUNT 56234 a 35308 c 35819 g 58686 t 1231 others
ORIGIN

Query Match 10.2%; Score 42.8; DB 2; Length 187278;
Best Local Similarity 57.5%; Pred. No. 0.38;
Matches 77; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

```
Qy      2 aaaaaaataactcggaaaagaaggagacgccgaaaattcgaaaggggaggggaaagcaaa 61
          | || ||  |   ||| ||||||||||| |  ||| |   | | | ||||| | | |  ||
Db  47386 AGAAGAAAGAAGAGGAGAAGAAGGAGAAGGAGAAGAGAAGGAGGAGGAGGAGGAGGAGAA 47445

Qy     62 gctgatggcggaggcccaggggaaagcaaagcaaatggcggaggccccgagcaagatcga 121
          |  || || | |||   ||| ||| |  ||| | | || | |||    |   ||||   |
Db  47446 GGAGAAGGAGAAGGAGAAGGAGAAGGAGAAGGAGAAGGAGAAGGAGAAGAAGAA 47505

Qy    122 atccatgaggaagt 135
              | || ||| |
Db  47506 GAAGAAGAAGAATT 47519
```

RESULT 12
AC025579/c
LOCUS AC025579 221104 bp DNA HTG 17-JUL-2001
DEFINITION Mus musculus clone RP23-286A4, WORKING DRAFT SEQUENCE, 4 unordered pieces.
ACCESSION AC025579
VERSION AC025579.23 GI:14787152
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE house mouse.
  ORGANISM Mus musculus
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 221104)
  AUTHORS Metzker,M.L., Lewis,L.R., Hume,J., Edwards,C., Harris,C., Dederich,D., Thomas,S., Okwuonu,G., Carlock,C., Garner,T., Addison,S., Pace,A., Williams,G., Bonnin,D., Brooks,A., Brown,J., Buhay,C., Bunac,C., Burkett,C., Chacko,J., Chen,G., Chen,Z., Cox,C., Davis,C., Delgado,O., Ding,Y., Dugan-Rocha,S., Fernandez,C., Ferraguto,D., Forcum-Tansey,J., Gill,R., Gorrell,J.H., Gunaratne,P., Haller,G., Hernandez,J., Hogues,M., Hosak,H., Hou,X., Huber,J., Jackson,L., Jia,Y., Kelly,J., Kelly,S., Kovar,C., Liu,J., Liu,W., Loulseged,H., Lozado,R.J., Martin,R., Massey,E., McLeod,M.P., Mei,G., Moore,S., Morgan,M., Morris,S., Neal,D., Nelson,A., Nguyen,R., Nguyen,N., Oguh,M., Parish,B., Perez,L., Reiter,D., Say,J., Shen,H., Vasquez,L., Watlington,S., Williamson,A., Wrensford,G., Zhou,X., Bouck,J., Hodgson,A., Muzny,D.M., Rives,M., Scherer,S., Sodergren,E., Weinstock,G., Worley,K. and Gibbs,R.
  TITLE Direct Submission
  JOURNAL Unpublished
REFERENCE 2 (bases 1 to 221104)
  AUTHORS Worley,K.C.
  TITLE Direct Submission
  JOURNAL Submitted (11-MAR-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

```
COMMENT     On Jul 17, 2001 this sequence version replaced gi:14547759.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: MABH
                Center clone name: RP23-286A4
            -------------- Summary Statistics
                Sequencing vector: M13; L08821
                Chemistry: Dye-primer Bodipy: 53% of reads
                Chemistry: Dye-terminator Big Dye: 47% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 222457 bases at least Q40
                Consensus quality: 223927 bases at least Q30
                Consensus quality: 224923 bases at least Q20
                Estimated insert size: 219940; sum-of-contigs estimation
                Quality coverage: 0x in Q20 bases; agarose-fp estimation
                Quality coverage: 7.9x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 4 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    67274: contig of 67274 bp in length
            *    67275    67374: gap of unknown length
            *    67375   132337: contig of 64963 bp in length
            *   132338   132437: gap of unknown length
            *   132438   188824: contig of 56387 bp in length
            *   188825   188924: gap of unknown length
            *   188925   221104: contig of 32180 bp in length.
FEATURES             Location/Qualifiers
     source          1. .221104
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="RP23-286A4"
BASE COUNT    57615 a  55079 c  54460 g  53650 t    300 others
ORIGIN


  Query Match             10.1%;  Score 42.4;  DB 2;  Length 221104;
  Best Local Similarity   55.4%;  Pred. No. 0.47;
  Matches   82;  Conservative    0;  Mismatches   66;  Indels    0;  Gaps    0;

Qy       5 aaaataactcggaaaagaaggagacgccgaaaattcgaaaggggaggggaaagcaaagct 64
           |||| ||   |   | |||| ||| |  || ||   ||| | | ||| ||| |  |||
Db 131331 AAAAAAAAAAGAGGAGGAAGAAGAGGAGGAGAAGGAGAAGGAGAAGGAGAAGGA 131272

Qy      65 gatggcggaggcccaggggaaagcaaagcaaatggcggaggccccgagcaagatcgaatc 124
           || || | |||   ||| ||| |  ||| | | || | |||    ||  ||||   |
```

```
Db 131271 GAAGGAGAAGGAGAAGGAGAAGGAGAAGGAGAAGGAGAAGGAGAAGAAGAAGAAGAAGAA 131212

Qy    125 catgaggaagtgggtcgtcgagcacaag 152
          | || |||| |       || | |||
Db 131211 GAAGAAGAAGAAGAAGAAGAAGAAGAAG 131184


RESULT 13
AC083834/c
LOCUS       AC083834   241714 bp    DNA             HTG       27-JUN-2001
DEFINITION  Mus musculus chromosome 11 clone RP23-51J10, WORKING DRAFT
            SEQUENCE, 5 unordered pieces.
ACCESSION   AC083834
VERSION     AC083834.20  GI:14547775
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 241714)
  AUTHORS   Metzker,M.L., Lewis,L.R., Hume,J., Edwards,C., Harris,C.,
            Dederich,D., Thomas,S., Okwuonu,G., Carlock,C., Garner,T.,
            Addison,S., Pace,A., Williams,G., Bonnin,D., Brooks,A., Brown,J.,
            Buhay,C., Bunac,C., Burkett,C., Chacko,J., Chen,G., Chen,Z.,
            Cox,C., Davis,C., Delgado,O., Ding,Y., Dugan-Rocha,S.,
            Fernandez,C., Ferraguto,D., Forcum-Tansey,J., Gill,R.,
            Gorrell,J.H., Gunaratne,P., Haller,G., Hernandez,J., Hogues,M.,
            Hosak,H., Hou,X., Huber,J., Jackson,L., Jia,Y., Kelly,J., Kelly,S.,
            Kovar,C., Liu,J., Liu,W., Loulseged,H., Lozado,R.J., Martin,R.,
            Massey,E., McLeod,M.P., Mei,G., Moore,S., Morgan,M., Morris,S.,
            Neal,D., Nelson,A., Nguyen,R., Nguyen,N., Oguh,M., Parish,B.,
            Perez,L., Reiter,D., Say,J., Shen,H., Vasquez,L., Watlington,S.,
            Williamson,A., Wrensford,G., Zhou,X., Bouck,J., Hodgson,A.,
            Muzny,D.M., Rives,M., Scherer,S., Sodergren,E., Weinstock,G.,
            Worley,K. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 241714)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-OCT-2000) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Jun 25, 2001 this sequence version replaced gi:13096022.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: MAPR
                Center clone name: RP23-51J10
            -------------- Summary Statistics
                Sequencing vector: M13; L08821
                Chemistry: Dye-primer Bodipy: 55% of reads
                Chemistry: Dye-terminator Big Dye: 45% of reads
                Assembly program: Phrap; version 0.990329
```

```
            Consensus quality: 239599 bases at least Q40
            Consensus quality: 241588 bases at least Q30
            Consensus quality: 243225 bases at least Q20
            Estimated insert size: 243199; sum-of-contigs estimation
            Quality coverage: 0x in Q20 bases; agarose-fp estimation
            Quality coverage: 8.2x in Q20 bases; sum-of-contigs estimation
          --------------
          * NOTE: Estimated insert size may differ from sequence length
          *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 5 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *        1    59312: contig of 59312 bp in length
          *    59313    59412: gap of unknown length
          *    59413   113658: contig of 54246 bp in length
          *   113659   113758: gap of unknown length
          *   113759   165679: contig of 51921 bp in length
          *   165680   165779: gap of unknown length
          *   165780   206158: contig of 40379 bp in length
          *   206159   206258: gap of unknown length
          *   206259   241714: contig of 35456 bp in length.
FEATURES             Location/Qualifiers
     source          1. .241714
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /chromosome="11"
                     /clone="RP23-51J10"
BASE COUNT    65862 a  55533 c  55480 g  64438 t    401 others
ORIGIN


  Query Match             10.1%;  Score 42.4;  DB 2;  Length 241714;
  Best Local Similarity   55.4%;  Pred. No. 0.47;
  Matches   82;  Conservative    0;  Mismatches   66;  Indels    0;  Gaps    0;

Qy      5 aaaataactcggaaaagaaggagacgccgaaaattcgaaaggggaggggaaagcaaagct 64
          |||| ||   |   | |||| ||| |  || ||   ||| | | ||| ||| |  |||
Db 217714 AAAAAAAAAAGAGGAGGAAGAAGAGGAGGAGAAGGAGAAGGAGAAGGAGAAGGAGAAGGA 217655

Qy     65 gatggcggaggcccaggggaaagcaaagcaaatggcggaggccccgagcaagatcgaatc 124
          || || | |||   ||| ||| |  ||| | | || | |||    ||  ||||   |
Db 217654 GAAGGAGAAGGAGAAGGAGAAGGAGAAGGAGAAGGAGAAGGAGAAGAAGAAGAAGAAGAA 217595

Qy    125 catgaggaagtgggtcgtcgagcacaag 152
           | || ||||  |       || | |||
Db 217594 GAAGAAGAAGAAGAAGAAGAAGAAGAAG 217567


RESULT  14
AC092197
LOCUS       AC092197   143977 bp    DNA             HTG       27-JUL-2001
DEFINITION  Homo sapiens chromosome 3q clone RP11-67H6, WORKING DRAFT SEQUENCE,
```

11 unordered pieces.
ACCESSION AC092197
VERSION AC092197.5 GI:15011619
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 143977)
AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstock,G. and Gibbs,R.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 143977)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (27-JUN-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT On Jul 25, 2001 this sequence version replaced gi:14971137.

```
            ---------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            ---------- Project Information
                Center project name: HCHC
                Center clone name: RP11-67H6
            ---------- Summary Statistics
                Sequencing vector: Plasmid; M77789
                Chemistry: Dye-terminator Big Dye: 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 144495 bases at least Q40
                Consensus quality: 146471 bases at least Q30
                Consensus quality: 148226 bases at least Q20
                Estimated insert size: 144773; sum-of-contigs estimation
                Quality coverage: 0x in Q20 bases; agarose-fp estimation
                Quality coverage: 8.8x in Q20 bases; sum-of-contigs estimation
            ----------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 11 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    24556: contig of 24556 bp in length
            *    24557    24656: gap of unknown length
            *    24657    48219: contig of 23563 bp in length
            *    48220    48319: gap of unknown length
            *    48320    69392: contig of 21073 bp in length
            *    69393    69492: gap of unknown length
            *    69493    88787: contig of 19295 bp in length
            *    88788    88887: gap of unknown length
            *    88888    99892: contig of 11005 bp in length
            *    99893    99992: gap of unknown length
            *    99993   111041: contig of 11049 bp in length
            *   111042   111141: gap of unknown length
            *   111142   118551: contig of 7410 bp in length
            *   118552   118651: gap of unknown length
            *   118652   128074: contig of 9423 bp in length
            *   128075   128174: gap of unknown length
            *   128175   135844: contig of 7670 bp in length
            *   135845   135944: gap of unknown length
            *   135945   141503: contig of 5559 bp in length
            *   141504   141603: gap of unknown length
            *   141604   143977: contig of 2374 bp in length.
FEATURES             Location/Qualifiers
     source          1. .143977
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="3q"
                     /clone="RP11-67H6"
BASE COUNT    43520 a  27330 c  27935 g  44191 t   1001 others
```

ORIGIN

Query Match 10.0%; Score 42; DB 2; Length 143977;
Best Local Similarity 56.5%; Pred. No. 0.63;
Matches 78; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

```
Qy      1 gaaaaaaataactcggaaaagaaggagacgccgaaaattcgaaaggggaggggaaagcaa 60
          | |  ||| ||   | | || || ||   |  |  |    || | |||||| ||    ||
Db  24385 GGAGGAAAGAAAAAGAAGAAAAAAGAAGAGAAGGGATGAAGAGAAGGGAGGAGAGGAGAA 24444

Qy     61 agctgatggcggaggcccaggggaaagcaaagcaaatggcggaggccccgagcaagatcg 120
          ||  ||||| ||||    ||||||  || ||||  || || ||| |    |||   |  |
Db  24445 AGAGGATGGAGGAGAGAGAGGGGAGAGAAAAGAGAAAGGGGGAAGGAGGGAGAGGGGACA 24504

Qy    121 aatccatgaggaagtggg 138
          ||   | |||| || | |
Db  24505 AAGGGAAGAGGGAGAGAG 24522
```

RESULT 15
AC007426/c
LOCUS AC007426 84199 bp DNA HTG 24-MAR-2000
DEFINITION Homo sapiens clone RP5-241A3, WORKING DRAFT SEQUENCE, 12 unordered pieces.
ACCESSION AC007426
VERSION AC007426.2 GI:7321581
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 84199)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens, clone RP5-241A3
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 84199)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M., Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A., Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A., Cooke,P., DeArellano,K., Depayre,E., Devon,K., Dewar,K., Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G., Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L., Karatas,A., Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J., Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B., Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P., Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A., Tesfaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A., Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
TITLE Direct Submission
JOURNAL Submitted (29-APR-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On Mar 24, 2000 this sequence version replaced gi:4713974.
All repeats were identified using RepeatMasker:

```
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L578
                Center clone name: 241_A_3
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 96% of reads
                Sequencing vector: Plasmid; n/a; %-0.f%% of reads
                3.5036496350365Chemistry: Dye-primer-amersham; 96% of reads
                Chemistry: Dye-terminator Big Dye; 4% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 76051 bases at least Q40
                Consensus quality: 79580 bases at least Q30
                Consensus quality: 81167 bases at least Q20
                Insert size: 83000; agarose-fp
                Insert size: 83099; sum-of-contigs
                Quality coverage: 4.1.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 12 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1147: contig of 1147 bp in length
            *     1148 1247: gap of      100 bp
            *     1248     2658: contig of 1411 bp in length
            *     2659 2758: gap of      100 bp
            *     2759     3896: contig of 1138 bp in length
            *     3897 3996: gap of      100 bp
            *     3997     5111: contig of 1115 bp in length
            *     5112 5211: gap of      100 bp
            *     5212     8025: contig of 2814 bp in length
            *     8026 8125: gap of      100 bp
            *     8126    10951: contig of 2826 bp in length
            *    10952 11051: gap of      100 bp
            *    11052    15725: contig of 4674 bp in length
            *    15726 15825: gap of      100 bp
            *    15826    21857: contig of 6032 bp in length
            *    21858 21957: gap of      100 bp
            *    21958    28985: contig of 7028 bp in length
            *    28986 29085: gap of      100 bp
            *    29086    42641: contig of 13556 bp in length
            *    42642 42741: gap of      100 bp
            *    42742    66388: contig of 23647 bp in length
            *    66389 66488: gap of      100 bp
            *    66489    84199: contig of 17711 bp in length.
FEATURES             Location/Qualifiers
     source          1. .84199
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
```

```
                     /clone="RP5-241A3"
                     /clone_lib="RPCI Human Male PAC"
     misc_feature    1. .1147
                     /note="assembly_fragment"
     misc_feature    1248. .2658
                     /note="assembly_fragment"
     misc_feature    2759. .3896
                     /note="assembly_fragment"
     misc_feature    3997. .5111
                     /note="assembly_fragment"
     misc_feature    5212. .8025
                     /note="assembly_fragment"
     misc_feature    8126. .10951
                     /note="assembly_fragment"
     misc_feature    11052. .15725
                     /note="assembly_fragment"
     misc_feature    15826. .21857
                     /note="assembly_fragment"
     misc_feature    21958. .28985
                     /note="assembly_fragment"
     misc_feature    29086. .42641
                     /note="assembly_fragment"
     misc_feature    42742. .66388
                     /note="assembly_fragment"
     misc_feature    66489. .84199
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:right"
BASE COUNT    24761 a  14984 c  15272 g  28081 t   1101 others
ORIGIN


  Query Match              9.9%;  Score 41.8;  DB 2;  Length 84199;
  Best Local Similarity   57.1%;  Pred. No. 0.75;
  Matches   76;  Conservative    0;  Mismatches   57;  Indels    0;  Gaps    0;

Qy      2 aaaaaaataactcggaaaagaaggagacgccgaaaattcgaaaggggaggggaaagcaaa 61
          | || || ||   | | |||||  ||| |  ||| |    | || | ||| ||| |  ||
Db  19791 AGAAGAAGAAGAAGAAGAAGAACAAGAAGAGGAAGAAGAAAGAGAGAAGGAGAAGGAGAA 19732

Qy     62 gctgatggcggaggcccaggggaaagcaaagcaaatggcggaggccccgagcaagatcga 121
          |  || || | |||   ||| ||| |  ||| | | || | |||    ||  ||||   |
Db  19731 GGAGAAGGAGAAGGAGAAGGAGAAGGAGAAGGAGAAGGAGAAGGAGAAGAAGAAGAAGAA 19672

Qy    122 atccatgaggaag 134
            | || ||||
Db  19671 GAAGAAGAAGAAG 19659


Search completed: February  7, 2002, 10:51:11
Job time: 8997 sec

                          GenCore version 4.5
                 Copyright (c) 1993 - 2000  Compugen Ltd.
```

OM nucleic - nucleic search, using sw model

Run on: February 7, 2002, 09:36:54 ; Search time 428.31 Seconds
(without alignments)
842.693 Million cell updates/sec

Title: US-09-394-745-5893
Perfect score: 421
Sequence: 1 gaaaaaaataactcggaaaa.....ccatgttggttcctgcatgc 421

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_1101:* — would like Document.
1: /SIDS2/gcgdata/geneseq/geneseqn/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description

	1	119.8	28.5	431	21	AAC36114	Arabidopsis thalia
c	2	39.4	9.4	162450	21	AAZ86967	Retinoblastoma bin
	3	38	9.0	244	22	AAI26833	Probe #16766 for g
	4	38	9.0	244	22	AAI55620	Probe #24306 used
	5	38	9.0	597	22	AAI17628	Probe #7561 for ge
	6	38	9.0	597	22	AAI42551	Probe #11237 used
	7	37.4	8.9	196	22	AAI54321	Probe #23007 used
	8	37.4	8.9	503	22	AAI41282	Probe #9968 used t
	9	35.6	8.5	49999	20	AAZ23891	Murine LOBO genomi
	10	35.6	8.5	49999	20	AAZ23896	Murine LOBO homolo
c	11	35.2	8.4	475	22	AAI11457	Probe #1390 for ge
c	12	35.2	8.4	475	22	AAI32728	Probe #1414 used t
c	13	35.2	8.4	475	22	AAI01373	Probe #1364 used t
c	14	35.2	8.4	512	22	AAI20671	Probe #10604 for g
c	15	35.2	8.4	512	22	AAI45882	Probe #14568 used
c	16	35.2	8.4	512	22	AAI06370	Probe #6361 used t
	17	34.4	8.2	232	22	AAI54322	Probe #23008 used
	18	34.4	8.2	559	22	AAI41283	Probe #9969 used t
	19	34.4	8.2	714	21	AAF13095	Aspergillus oryzae
	20	34.2	8.1	2643	14	AAQ39212	CENP-B cDNA. Homo
	21	34.2	8.1	3717	21	AAA64660	DNA encoding centr
c	22	34.2	8.1	49999	20	AAZ23895	Murine LOBO homolo
	23	34	8.1	208	22	AAI52685	Probe #21371 used
c	24	34	8.1	249	21	AAC05635	Human secreted pro
c	25	34	8.1	249	21	AAC05637	Human secreted pro
	26	34	8.1	347	22	AAI23316	Probe #13249 for g
	27	34	8.1	347	22	AAI48637	Probe #17323 used
	28	34	8.1	488	22	AAI14106	Probe #4039 for ge
	29	34	8.1	488	22	AAI35487	Probe #4173 used t
	30	34	8.1	567	22	AAI39603	Probe #8289 used t
	31	34	8.1	3756	18	AAT72684	Sugar biosynthesis
c	32	33.8	8.0	33030	22	AAF29337	Atopy related gene
c	33	33.6	8.0	272	22	AAI21218	Probe #11151 for g
c	34	33.6	8.0	272	22	AAI46489	Probe #15175 used
c	35	33.6	8.0	272	22	AAI06928	Probe #6919 used t
	36	33.6	8.0	989	21	AAA02539	Human colon cancer
	37	33.4	7.9	2175	20	AAX24906	Mouse elongation f
	38	33.4	7.9	8942	22	AAF29338	Atopy related gene
	39	33.2	7.9	349	22	AAI14995	Probe #4928 for ge
	40	33.2	7.9	349	22	AAI36341	Probe #5027 used t
	41	33.2	7.9	349	22	AAI04761	Probe #4752 used t
	42	33.2	7.9	385	21	AAC02428	Human secreted pro
c	43	33.2	7.9	1763	12	AAQ10025	Encodes a portion
c	44	33.2	7.9	1763	14	AAQ47732	Coding sequence fo
c	45	33	7.8	626	22	AAI22123	Probe #12056 for g

ALIGNMENTS

RESULT 1
AAC36114
ID AAC36114 standard; DNA; 431 BP.
XX
AC AAC36114;
XX
DT 17-OCT-2000 (first entry)

XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 12597.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.

PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.

PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.

```
PR   14-OCT-1999;   99US-0159330.
PR   14-OCT-1999;   99US-0159331.
PR   14-OCT-1999;   99US-0159637.
PR   14-OCT-1999;   99US-0159638.
PR   18-OCT-1999;   99US-0159584.
PR   21-OCT-1999;   99US-0160741.
PR   21-OCT-1999;   99US-0160767.
PR   21-OCT-1999;   99US-0160768.
PR   21-OCT-1999;   99US-0160770.
PR   21-OCT-1999;   99US-0160814.
PR   21-OCT-1999;   99US-0160815.
PR   22-OCT-1999;   99US-0160980.
PR   22-OCT-1999;   99US-0160981.
PR   22-OCT-1999;   99US-0160989.
PR   25-OCT-1999;   99US-0161404.
PR   25-OCT-1999;   99US-0161405.
PR   25-OCT-1999;   99US-0161406.
PR   26-OCT-1999;   99US-0161359.
PR   26-OCT-1999;   99US-0161360.
PR   26-OCT-1999;   99US-0161361.
PR   28-OCT-1999;   99US-0161920.
PR   28-OCT-1999;   99US-0161992.
PR   28-OCT-1999;   99US-0161993.
PR   29-OCT-1999;   99US-0162142.

  Query Match             28.5%;  Score 119.8;  DB 21;  Length 431;
  Best Local Similarity   67.3%;  Pred. No. 6e-29;
  Matches  169;  Conservative    0;  Mismatches   82;  Indels    0;  Gaps    0;

Qy      95 aatggcggaggccccgagcaagatcgaatccatgaggaagtgggtcgtcgagcacaagct 154
           |||||||||  |   ||  ||  | | |   || ||| | ||| || |||| || |||||
Db      31 aatggcggaaccaaagacaaaagttgcagaaatcagggaatggatcatcgaacataagct 90

Qy     155 ccgagccgtaggttgcctctggctaggtgggatcagcagttcgatcgcctacaactggtc 214
            ||  |||| |||||| | |||||| |||| |||    |||| || || || || |||||
Db      91 tcgtaccgttggttgcttatggctaagtggtatctctggttcaattgcttataattggtc 150

Qy     215 gcggcccaatatgaagcctagcgtcaagatcatccacgcaaggttgcatgctcaagctct 274
               ||    |||||  | || ||||  |||||||||||| |||||||||||||| || ||
Db     151 taaacctgccatgaaaaccagtgtcagaatcatccacgctaggttgcatgctcaggcgct 210

Qy     275 aaccctggctgcattagttggttctgcatgcgtggagtactatgaccagaagtatggttc 334
            ||  | || ||  | | |||  | ||    |||||||||||||| || || | |||  |
Db     211 gacattagccgctctggctggagcagctgcagtggagtactatgatcacaaatctggagc 270

Qy     335 ttctgggccaa 345
             |||  | ||
Db     271 cactgatcgaa 281


RESULT   2
AAZ86967/c
ID   AAZ86967 standard; DNA; 162450 BP.
XX
AC   AAZ86967;
XX
DT   16-MAY-2000  (first entry)
```

XX
DE Retinoblastoma binding protein-7 genomic DNA sequence.
XX
KW RBP-7; retinoblastoma binding protein-7; abnormal cell proliferation;
KW diagnosis; therapy; cell differentiation; thyroid hyperplasia; psoriasis;
KW benign prostate hypertrophy; cancer; sarcoma; neoplasm; leukaemia;
KW lymphoma; ds.
XX
OS Homo sapiens.
XX
PN WO200000607-A1.
XX
PD 06-JAN-2000.
XX
PF 30-JUN-1999; 99WO-IB01242.
XX
PR 30-JUN-1998; 98US-0091315.
PR 10-DEC-1998; 98US-0111909.
XX
PA (GEST) GENSET.
XX
PI Bougueleret L;
XX
DR WPI; 2000-117170/10.
XX
PT Novel nucleic acid and polymorphic markers used for diagnosis of
PT diseases, especially those involving abnormal cell proliferation and
PT differentiation -
XX
PS Claim 1; Page 118-163; 223pp; English.
XX
CC This sequence represents the retinoblastoma binding protein-7 (RBP-7)
CC genomic sequence of the invention. The RBP-7 coding sequence and
CC regulatory sequences are useful for the recombinant production of the
CC protein and for expressing heterologous nucleic acids. Primers and
CC probes derived from the RBP-7 nucleotide sequence (e.g. AAZ87035-Z87099)
CC are useful for DNA amplification and detection methods. RBP-7 biallelic
CC markers (see AAZ86993-Z87034) are useful for diagnosis of disease
CC related to alteration in the regulation or in the coding regions of the
CC RBP-7 gene and for prognosis/diagnosis of an eventual treatment with
CC therapeutic agents, especially agents acting on pathologies involving
CC abnormal cell proliferation and/or differentiation, these include
CC thyroid hyperplasia, psoriasis, benign prostate hypertrophy, cancers,
CC including breast cancer, sarcomas and other neoplasms, bladder cancer,
CC colon cancer, lung cancer, prostate cancer, various leukaemias, and
CC lymphomas. RBP-7 antibodies are useful as diagnostic agents.
XX
SQ Sequence 162450 BP; 45465 A; 30661 C; 32637 G; 53673 T; 14 other;

Query Match 9.4%; Score 39.4; DB 21; Length 162450;
Best Local Similarity 55.5%; Pred. No. 0.089;
Matches 76; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

Qy 1 gaaaaaaataactcggaaaagaaggagacgccgaaaattcgaaaggggaggggaaagcaa 60
Db 26250 GCAAAAAAAAACAAAGAAACAAGGAAAAGGAAGGGAAGGGGAAGGGAAAGGGGAAGGGGA 26191

```
Qy     61 agctgatggcggaggcccaggggaaagcaaagcaaatggcggaggccccgagcaagatcg 120
          |   ||||| |  ||   || | ||||  | | |   |  | |||     || |||
Db  26190 AAGAGATGGGGAGGGGGAAGTGTAAAGGGAGGGAGGGGAGGAAGGAAGGAAGGAAGGAAA 26131

Qy    121 aatccatgaggaagtgg 137
          ||   |   ||||||||
Db  26130 AAGGGAAAGGGAAGTGG 26114
```

RESULT 3
AAI26833
ID AAI26833 standard; DNA; 244 BP.
XX
AC AAI26833;
XX
DT 12-OCT-2001 (first entry)
XX
DE Probe #16766 for gene expression analysis in human cervical cell sample.
XX
KW Probe; human; microarray; gene expression; cervical epithelial cell;
KW cervical cancer; ss.
XX
OS Homo sapiens.
XX
PN WO200157278-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00670.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488901/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human cervical epithelial cells -
XX
PS Claim 25; SEQ ID No 16766; 487pp; English.
XX
CC The present invention relates to human single exon nucleic acid probes
CC (SENP). The present sequence is one such probe. The SENPs are derived
CC from human HeLa cells. The SENPs can be used to produce a single exon
CC microarray, which can be used for measuring human gene expression in a
CC sample derived from human cervical epithelial cells. By measuring gene
CC expression, the probes are therefore useful in grading and/or staging
CC of diseases of the cervix, notably cervical cancer.

CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 244 BP; 101 A; 25 C; 81 G; 37 T; 0 other;

Query Match 9.0%; Score 38; DB 22; Length 244;
Best Local Similarity 60.8%; Pred. No. 0.014;
Matches 62; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

```
Qy       4 aaaaataactcggaaaagaaggagacgccgaaaattcgaaaggggaggggaaagcaaagc 63
           ||||| ||    |  ||||||||||| |  ||| |   | ||  ||| ||| ||| ||||
Db     129 aaaaagaaagagagaaagaaggagaaggagaagaagaggaacaggaaggggaagaaaaga 188

Qy      64 tgatggcggaggcccaggggaaagcaaagcaaatggcggagg 105
             || |  | || |  || ||||     || |   || ||| |
Db     189 agaagaagaagacgaagaggaagaaggaggaggaggaggaag 230
```

RESULT 4
AAI55620
ID AAI55620 standard; DNA; 244 BP.
XX
AC AAI55620;
XX
DT 17-OCT-2001 (first entry)
XX
DE Probe #24306 used to measure gene expression in human placenta sample.
XX
KW Probe; microarray; human; placenta; antenatal diagnosis;
KW genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200157272-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00663.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488897/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human placenta -

XX
PS Claim 25; SEQ ID No 24306; 654pp; English.
XX
CC The present invention relates to single exon nucleic acid probes (SENP).
CC The present sequence is one such probe. The probes are useful for
CC producing a microarray for predicting, measuring and displaying gene
CC expression in samples derived from human placenta. The probes are useful
CC for antenatal diagnosis of human genetic disorders.
XX
SQ Sequence 244 BP; 101 A; 25 C; 81 G; 37 T; 0 other;

Query Match 9.0%; Score 38; DB 22; Length 244;
Best Local Similarity 60.8%; Pred. No. 0.014;
Matches 62; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

```
Qy       4 aaaaataactcggaaaagaaggagacgccgaaaattcgaaaggggagggggaaagcaaagc 63
           ||||| ||   |  ||||||||||| |  ||| |   | ||  ||| ||| ||| ||||
Db     129 aaaaagaaagagagaaagaaggagaaggagaagaagaggaacaggaaggggaagaaaaga 188

Qy      64 tgatggcggaggcccaggggaaagcaaagcaaatggcggagg 105
              || |  | || |  || ||||     || |   || ||| |
Db     189 agaagaagaagacgaagaggaagaaggaggaggaggaggaag 230
```

RESULT 5
AAI17628
ID AAI17628 standard; DNA; 597 BP.
XX
AC AAI17628;
XX
DT 12-OCT-2001 (first entry)
XX
DE Probe #7561 for gene expression analysis in human cervical cell sample.
XX
KW Probe; human; microarray; gene expression; cervical epithelial cell;
KW cervical cancer; ss.
XX
OS Homo sapiens.
XX
PN WO200157278-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00670.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX
DR WPI; 2001-488901/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human cervical epithelial cells -
XX
PS Claim 25; SEQ ID No 7561; 487pp; English.
XX
CC The present invention relates to human single exon nucleic acid probes
CC (SENP). The present sequence is one such probe. The SENPs are derived
CC from human HeLa cells. The SENPs can be used to produce a single exon
CC microarray, which can be used for measuring human gene expression in a
CC sample derived from human cervical epithelial cells. By measuring gene
CC expression, the probes are therefore useful in grading and/or staging
CC of diseases of the cervix, notably cervical cancer.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 597 BP; 228 A; 67 C; 170 G; 132 T; 0 other;

Query Match 9.0%; Score 38; DB 22; Length 597;
Best Local Similarity 60.8%; Pred. No. 0.021;
Matches 62; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

```
Qy      4 aaaaataactcggaaaagaaggagacgccgaaaattcgaaaggggaggggaaagcaaagc 63
          ||||| ||   |  ||||||||||| |  ||| |   | ||  ||| ||| ||| ||||
Db    389 aaaaagaaagagagaaagaaggagaaggagaagaagaggaacaggaaggggaagaaaaga 448

Qy     64 tgatggcggaggcccaggggaaagcaaagcaaatggcggagg 105
           || |  | || |  || ||||     || |   || ||| |
Db    449 agaagaagaagacgaagaggaagaaggaggaggaggaggaag 490
```

RESULT 6
AAI42551
ID AAI42551 standard; DNA; 597 BP.
XX
AC AAI42551;
XX
DT 17-OCT-2001 (first entry)
XX
DE Probe #11237 used to measure gene expression in human placenta sample.
XX
KW Probe; microarray; human; placenta; antenatal diagnosis;
KW genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200157272-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00663.
XX
PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488897/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human placenta -
XX
PS Claim 25; SEQ ID No 11237; 654pp; English.
XX
CC The present invention relates to single exon nucleic acid probes (SENP).
CC The present sequence is one such probe. The probes are useful for
CC producing a microarray for predicting, measuring and displaying gene
CC expression in samples derived from human placenta. The probes are useful
CC for antenatal diagnosis of human genetic disorders.
XX
SQ Sequence 597 BP; 228 A; 67 C; 170 G; 132 T; 0 other;

  Query Match   9.0%; Score 38; DB 22; Length 597;
  Best Local Similarity   60.8%; Pred. No. 0.021;
  Matches   62; Conservative   0; Mismatches   40; Indels   0; Gaps   0;

Qy   4 aaaaataactcggaaaagaaggagacgccgaaaattcgaaaggggaggggaaagcaaagc 63
           ||||| || | |||||||||||| | ||| | | || ||| ||| ||| ||||
Db   389 aaaaagaaagagagaaagaaggagaaggagaagaagaggaacaggaaggggaagaaaaga 448

Qy   64 tgatggcggaggcccaggggaaagcaaagcaaatggcggagg 105
              || | | || | || |||| || | || ||| |
Db   449 agaagaagaagacgaagaggaagaaggaggaggaggaggaag 490

RESULT 7
AAI54321
ID AAI54321 standard; DNA; 196 BP.
XX
AC AAI54321;
XX
DT 17-OCT-2001 (first entry)
XX
DE Probe #23007 used to measure gene expression in human placenta sample.
XX
KW Probe; microarray; human; placenta; antenatal diagnosis;
KW genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200157272-A2.
XX

PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00663.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488897/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human placenta -
XX
PS Claim 25; SEQ ID No 23007; 654pp; English.
XX
CC The present invention relates to single exon nucleic acid probes (SENP).
CC The present sequence is one such probe. The probes are useful for
CC producing a microarray for predicting, measuring and displaying gene
CC expression in samples derived from human placenta. The probes are useful
CC for antenatal diagnosis of human genetic disorders.
XX
SQ Sequence 196 BP; 102 A; 8 C; 85 G; 1 T; 0 other;

Query Match 8.9%; Score 37.4; DB 22; Length 196;
Best Local Similarity 53.0%; Pred. No. 0.019;
Matches 80; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

```
Qy       2 aaaaaaataactcggaaaagaaggagacgccgaaaattcgaaaggggaggggaaagcaaa 61
           ||||  |  |   | |  |||||||||| |  ||  |   | | | ||||| ||||    |
Db      32 aaaagcagcagcagcaggagaaggagaagaggaggaggaggaggaggaggagaaaaagga 91

Qy      62 gctgatggcggaggcccaggggaaagcaaagcaaatggcggaggccccgagcaagatcga 121
           |  |  || ||||    ||| | | |   || ||| || |||||    |   ||||   |
Db      92 ggaggaggaggagaaggaggaggaggaggagaaaaaggaggaggagaaggagaagaagaa 151

Qy     122 atccatgaggaagtgggtcgtcgagcacaag 152
               | || ||||  |       || | |||
Db     152 ggagaagaagaagaagaagaggaagaagaag 182
```

RESULT 8
AAI41282
ID AAI41282 standard; DNA; 503 BP.
XX
AC AAI41282;
XX
DT 17-OCT-2001 (first entry)
XX

DE Probe #9968 used to measure gene expression in human placenta sample.
XX
KW Probe; microarray; human; placenta; antenatal diagnosis;
KW genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200157272-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00663.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488897/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human placenta -
XX
PS Claim 25; SEQ ID No 9968; 654pp; English.
XX
CC The present invention relates to single exon nucleic acid probes (SENP).
CC The present sequence is one such probe. The probes are useful for
CC producing a microarray for predicting, measuring and displaying gene
CC expression in samples derived from human placenta. The probes are useful
CC for antenatal diagnosis of human genetic disorders.
XX
SQ Sequence 503 BP; 207 A; 59 C; 143 G; 94 T; 0 other;

Query Match 8.9%; Score 37.4; DB 22; Length 503;
Best Local Similarity 53.0%; Pred. No. 0.03;
Matches 80; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

```
Qy      2 aaaaaaataactcggaaaagaaggagacgccgaaaattcgaaaggggaggggaaagcaaa 61
          ||||  |  |   | |  |||||||||  |  ||  |   | | | |||||  ||||    |
Db    306 aaaagcagcagcagcaggagaaggagaagaggaggaggaggaggaggaggagaaaaagga 365

Qy     62 gctgatggcggaggcccaggggaaagcaaagcaaatggcggaggccccgagcaagatcga 121
          |  |  || ||||    ||| | | |   || ||| || |||||    |   ||||   |
Db    366 ggaggaggaggagaaggaggaggaggaggagaaaaaggaggaggagaaggagaagaagaa 425

Qy    122 atccatgaggaagtgggtcgtcgagcacaag 152
              | || ||||  |       || | |||
Db    426 ggagaagaagaagaagaagaggaagaagaag 456
```

RESULT 9
AAZ23891
ID AAZ23891 standard; DNA; 49999 BP.
XX
AC AAZ23891;
XX
DT 25-JAN-2000 (first entry)
XX
DE Murine LOBO genomic DNA fragment 1.
XX
KW LOBO; long bones; bone development; bone extension; skull; osteopathic;
KW diagnostic; pharmaceutical; gene therapy; transgenic animal; disease;
KW spondyloepiphysal dysplasia; achondroplasia; murine; ds.
XX
OS Mus musculus.
XX
PN WO9950284-A2.
XX
PD 07-OCT-1999.
XX
PF 26-MAR-1999; 99WO-EP02055.
XX
PR 27-MAR-1998; 98DE-1013799.
XX
PA (ROSE/) ROSENTHAL A.
XX
PI Rosenthal A, Rump A, Hess J, Aigner T, Wirth T;
XX
DR WPI; 1999-601320/51.
XX
PT Nucleic acids encoding proteins which influence bone development,
PT useful for treating and studying bone disorders -
XX
PS Example 3; Page 69-97; 391pp; German.
XX
CC This invention describes novel nucleic acids (I; designated LOBO (long
CC bones)) encoding proteins influencing bone development in mammals. The
CC proteins of the invention reduce and/or inactivate bone extension (i.e.
CC development), with exception of the skull and have osteopathic activity.
CC The nucleic acid molecules, proteins and antibodies can be used in
CC diagnostic or pharmaceutical compounds e.g. for gene therapy. The methods
CC and nucleic acid molecules, etc. are useful for production of transgenic
CC animals, especially a transgenic mouse for the study of diseases
CC associated with bone development, e.g. spondyloepiphysal dysplasia and
CC achondroplasia. This sequence encodes the murine LOBO protein described
CC in the method of the invention.
XX
SQ Sequence 49999 BP; 13210 A; 11814 C; 10825 G; 14150 T; 0 other;

Query Match 8.5%; Score 35.6; DB 20; Length 49999;
Best Local Similarity 53.6%; Pred. No. 0.89;
Matches 74; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

Qy 15 ggaaaagaaggagacgccgaaaattcgaaaggggaggggaaagcaaagctgatggcggag 74

```
Db    7611 gaagaagaaggagacggagagaagaagaaggagaaggaaaaagagaagaagaagaaggag 7670

Qy      75 gcccaggggaaagcaaagcaaatggcggaggccccgagcaagatcgaatccatgaggaag 134
           ||    |     || | | || |||||    ||   |||   |    | || ||||
Db    7671 aaggagaaagagaaggagaagaaggaggaggaggagaaggagaagaagaagaagaagaag 7730

Qy     135 tgggtcgtcgagcacaag 152
             |       || | |||
Db    7731 aagaagaagaagaagaag 7748


RESULT  10
AAZ23896
ID   AAZ23896 standard; DNA; 49999 BP.
XX
AC   AAZ23896;
XX
DT   25-JAN-2000  (first entry)
XX
DE   Murine LOBO homologue genomic DNA fragment 2.
XX
KW   LOBO; long bones; bone development; bone extension; skull; osteopathic;
KW   diagnostic; pharmaceutical; gene therapy; transgenic animal; disease;
KW   spondyloepiphysal dysplasia; achondroplasia; murine; ds.
XX
OS   Mus musculus.
XX
PN   WO9950284-A2.
XX
PD   07-OCT-1999.
XX
PF   26-MAR-1999;   99WO-EP02055.
XX
PR   27-MAR-1998;   98DE-1013799.
XX
PA   (ROSE/) ROSENTHAL A.
XX
PI   Rosenthal A,  Rump A,  Hess J,  Aigner T,  Wirth T;
XX
DR   WPI; 1999-601320/51.
XX
PT   Nucleic acids encoding proteins which influence bone development,
PT   useful for treating and studying bone disorders -
XX
PS   Example 3; Page 161-189; 391pp; German.
XX
CC   This invention describes novel nucleic acids (I; designated LOBO (long
CC   bones)) encoding proteins influencing bone development in mammals. The
CC   proteins of the invention reduce and/or inactivate bone extension (i.e.
CC   development), with exception of the skull and have osteopathic activity.
CC   The nucleic acid molecules, proteins and antibodies can be used in
CC   diagnostic or pharmaceutical compounds e.g. for gene therapy. The methods
CC   and nucleic acid molecules, etc. are useful for production of transgenic
CC   animals, especially a transgenic mouse for the study of diseases
CC   associated with bone development, e.g. spondyloepiphysal dysplasia and
CC   achondroplasia. This sequence encodes the murine LOBO protein described
CC   in the method of the invention.
```

XX
SQ Sequence 49999 BP; 13135 A; 11787 C; 10868 G; 14209 T; 0 other;

Query Match 8.5%; Score 35.6; DB 20; Length 49999;
Best Local Similarity 53.6%; Pred. No. 0.89;
Matches 74; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

```
Qy     15 ggaaaagaaggagacgccgaaaattcgaaaggggaggggaaagcaaagctgatggcggag 74
          | | ||||||||||||  || ||   ||| | | |||  ||||  |||  || |  ||||
Db   9596 gaagaagaaggagacggagagaagaagaaggagaaggaaaaagagaagaagaagaaggag 9655

Qy     75 gcccaggggaaagcaaagcaaatggcggaggccccgagcaagatcgaatccatgaggaag 134
             ||    |     || | | || |||||    ||   |||   |    | || ||||
Db   9656 aaggagaaagagaaggagaagaaggaggaggaggagaaggagaagaagaagaagaagaag 9715

Qy    135 tgggtcgtcgagcacaag 152
            |       || | |||
Db   9716 aagaagaagaagaagaag 9733
```

RESULT 11
AAI11457/c
ID AAI11457 standard; DNA; 475 BP.
XX
AC AAI11457;
XX
DT 12-OCT-2001 (first entry)
XX
DE Probe #1390 for gene expression analysis in human cervical cell sample.
XX
KW Probe; human; microarray; gene expression; cervical epithelial cell;
KW cervical cancer; ss.
XX
OS Homo sapiens.
XX
PN WO200157278-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00670.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488901/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for

PT analyzing gene expression in human cervical epithelial cells -
XX
PS Claim 25; SEQ ID No 1390; 487pp; English.
XX
CC The present invention relates to human single exon nucleic acid probes
CC (SENP). The present sequence is one such probe. The SENPs are derived
CC from human HeLa cells. The SENPs can be used to produce a single exon
CC microarray, which can be used for measuring human gene expression in a
CC sample derived from human cervical epithelial cells. By measuring gene
CC expression, the probes are therefore useful in grading and/or staging
CC of diseases of the cervix, notably cervical cancer.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 475 BP; 24 A; 201 C; 30 G; 220 T; 0 other;

Query Match 8.4%; Score 35.2; DB 22; Length 475;
Best Local Similarity 52.0%; Pred. No. 0.15;
Matches 79; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

```
Qy       1 gaaaaaaataactcggaaaagaaggagacgccgaaaattcgaaaggggaggggaaagcaa 60
           ||| ||    |   | |  ||| |  || |   |  ||  | | | | ||| | | |
Db     314 GAAGAAGGAGAGAAGAAGGAGAGGAGGAGGAGCAGGATGAGGAGGAGCAGGAGGAGGAGG 255

Qy      61 agctgatggcggaggcccaggggaaagcaaagcaaatggcggaggccccgagcaagatcg 120
           ||| |  || |||||  |||| ||| |   ||||   |  |||||    |   ||||
Db     254 AGCAGGAGGAGGAGGAGCAGGAGAAGGAGGAGCAGGAGAAGGAGGAGGGGGAGAAGAAGA 195

Qy     121 aatccatgaggaagtgggtcgtcgagcacaag 152
           |    | || ||||  ||     ||| | |||
Db     194 AGGAGAAGAAGAAGGAGGAGAAGGAGGAGAAG 163
```

RESULT 12
AAI32728/c
ID AAI32728 standard; DNA; 475 BP.
XX
AC AAI32728;
XX
DT 17-OCT-2001 (first entry)
XX
DE Probe #1414 used to measure gene expression in human placenta sample.
XX
KW Probe; microarray; human; placenta; antenatal diagnosis;
KW genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200157272-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00663.
XX
PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488897/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human placenta -
XX
PS Claim 25; SEQ ID No 1414; 654pp; English.
XX
CC The present invention relates to single exon nucleic acid probes (SENP).
CC The present sequence is one such probe. The probes are useful for
CC producing a microarray for predicting, measuring and displaying gene
CC expression in samples derived from human placenta. The probes are useful
CC for antenatal diagnosis of human genetic disorders.
XX
SQ Sequence 475 BP; 24 A; 201 C; 30 G; 220 T; 0 other;

  Query Match   8.4%; Score 35.2; DB 22; Length 475;
  Best Local Similarity   52.0%; Pred. No. 0.15;
  Matches   79; Conservative   0; Mismatches   73; Indels   0; Gaps   0;

Qy   1 gaaaaaaataactcggaaaagaaggagacgccgaaaattcgaaaggggaggggaaagcaa 60
          ||| || | | | ||| | || | | || | | | | ||| | | |
Db   314 GAAGAAGGAGAGAAGAAGGAGAGGAGGAGGAGCAGGATGAGGAGGAGCAGGAGGAGGAGG 255

Qy   61 agctgatggcggaggcccaggggaaagcaaagcaaatggcggaggccccgagcaagatcg 120
          ||| | || ||||| |||| ||| | |||| | ||||| | ||||
Db   254 AGCAGGAGGAGGAGGAGCAGGAGAAGGAGGAGCAGGAGAAGGAGGAGGGGGAGAAGAAGA 195

Qy   121 aatccatgaggaagtgggtcgtcgagcacaag 152
          | | || |||| || ||| | |||
Db   194 AGGAGAAGAAGAAGGAGGAGAAGGAGGAGAAG 163

RESULT 13
AAI01373/c
ID AAI01373 standard; DNA; 475 BP.
XX
AC AAI01373;
XX
DT 09-OCT-2001 (first entry)
XX
DE Probe #1364 used to measure gene expression in human breast sample.
XX
KW Probe; human; breast disease; breast cancer; development disorder; ss;
KW inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX

OS Homo sapiens.
XX
PN WO200157270-A2.
XX
PD 09-AUG-2001.
XX
PF 29-JAN-2001; 2001WO-US00661.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-476286/51.
XX
PT Novel single exon nucleic acid probe used to measuring gene expression
PT in a human breast -
XX
PS Claim 25; SEQ ID No 1364; 322pp; English.
XX
CC The present invention relates to novel single exon nucleic acid probes.
CC The present sequence is one such probe. The probes are useful for
CC measuring human gene expression in a human breast sample, where the probe
CC hybridises at high stringency to a nucleic acid expressed in the human
CC breast. The probes are useful for predicting, diagnosing, grading,
CC staging, monitoring and prognosing diseases of the human breast,
CC particularly those diseases with polygenic aetiology. The diseases
CC include: breast cancer, disorders of development, inflammatory diseases
CC of the breast, fibrocystic changes, proliferative breast disease and
CC non-carcinoma tumours.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 475 BP; 24 A; 201 C; 30 G; 220 T; 0 other;

Query Match 8.4%; Score 35.2; DB 22; Length 475;
Best Local Similarity 52.0%; Pred. No. 0.15;
Matches 79; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

```
Qy      1 gaaaaaaataactcggaaaagaaggagacgccgaaaattcgaaaggggaggggaaagcaa 60
          ||| ||    |   | |  ||| |  || |   |  ||  | | | | ||| | | |
Db    314 GAAGAAGGAGAGAAGAAGGAGAGGAGGAGGAGCAGGATGAGGAGGAGCAGGAGGAGGAGG 255

Qy     61 agctgatggcggaggcccaggggaaagcaaagcaaatggcggaggccccgagcaagatcg 120
          ||| |  || |||||  |||| ||| |   ||||   |  |||||    |   ||||
Db    254 AGCAGGAGGAGGAGGAGCAGGAGAAGGAGGAGCAGGAGAAGGAGGAGGGGGAGAAGAAGA 195

Qy    121 aatccatgaggaagtgggtcgtcgagcacaag 152
```

```
          |    | || ||||  ||     ||| | |||
Db    194 AGGAGAAGAAGAAGGAGGAGAAGGAGGAGAAG 163


RESULT  14
AAI20671/c
ID   AAI20671 standard; DNA; 512 BP.
XX
AC   AAI20671;
XX
DT   12-OCT-2001  (first entry)
XX
DE   Probe #10604 for gene expression analysis in human cervical cell sample.
XX
KW   Probe; human; microarray; gene expression; cervical epithelial cell;
KW   cervical cancer; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157278-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00670.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-488901/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for
PT   analyzing gene expression in human cervical epithelial cells -
XX
PS   Claim 25; SEQ ID No 10604; 487pp; English.
XX
CC   The present invention relates to human single exon nucleic acid probes
CC   (SENP). The present sequence is one such probe. The SENPs are derived
CC   from human HeLa cells. The SENPs can be used to produce a single exon
CC   microarray, which can be used for measuring human gene expression in a
CC   sample derived from human cervical epithelial cells. By measuring gene
CC   expression, the probes are therefore useful in grading and/or staging
CC   of diseases of the cervix, notably cervical cancer.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 512 BP; 18 A; 231 C; 28 G; 235 T; 0 other;
```

Query Match 8.4%; Score 35.2; DB 22; Length 512;
Best Local Similarity 52.0%; Pred. No. 0.15;
Matches 79; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

```
Qy       1 gaaaaaaataactcggaaaagaaggagacgccgaaaattcgaaaggggagggggaaagcaa 60
           ||| ||    |   | |  ||| |  || |   |  ||  | | | | ||| | | |
Db     180 GAAGAAGGAGAGAAGAAGGAGAGGAGGAGGAGCAGGATGAGGAGGAGCAGGAGGAGGAGG 121

Qy      61 agctgatggcggaggcccaggggaaagcaaagcaaatggcggaggccccgagcaagatcg 120
           ||| |  || |||||  |||| ||| |   ||||   |  |||||    |   ||||
Db     120 AGCAGGAGGAGGAGGAGCAGGAGAAGGAGGAGCAGGAGAAGGAGGAGGGGGGAGAAGAAGA 61

Qy     121 aatccatgaggaagtgggtcgtcgagcacaag 152
           |    | || ||||  ||     ||| | |||
Db      60 AGGAGAAGAAGAAGGAGGAGAAGGAGGAGAAG 29
```

RESULT 15
AAI45882/c
ID AAI45882 standard; DNA; 512 BP.
XX
AC AAI45882;
XX
DT 17-OCT-2001 (first entry)
XX
DE Probe #14568 used to measure gene expression in human placenta sample.
XX
KW Probe; microarray; human; placenta; antenatal diagnosis;
KW genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200157272-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00663.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488897/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human placenta -
XX
PS Claim 25; SEQ ID No 14568; 654pp; English.

```
XX
CC   The present invention relates to single exon nucleic acid probes (SENP).
CC   The present sequence is one such probe. The probes are useful for
CC   producing a microarray for predicting, measuring and displaying gene
CC   expression in samples derived from human placenta. The probes are useful
CC   for antenatal diagnosis of human genetic disorders.
XX
SQ   Sequence 512 BP; 18 A; 231 C; 28 G; 235 T; 0 other;


  Query Match             8.4%;  Score 35.2;  DB 22;  Length 512;
  Best Local Similarity   52.0%;  Pred. No. 0.15;
  Matches   79;  Conservative    0;  Mismatches   73;  Indels    0;  Gaps    0;

Qy      1 gaaaaaaataactcggaaaagaaggagacgccgaaaattcgaaaggggaggggaaagcaa 60
          ||| ||    |   | |  ||| |  || |   |  ||  | | | | ||| | | |
Db    180 GAAGAAGGAGAGAAGAAGGAGAGGAGGAGGAGCAGGATGAGGAGGAGCAGGAGGAGGAGG 121

Qy     61 agctgatggcggaggcccaggggaaagcaaagcaaatggcggaggccccgagcaagatcg 120
          ||| |  || |||||  |||| ||| |   |||| . |  |||||    |   ||||
Db    120 AGCAGGAGGAGGAGGAGCAGGAGAAGGAGGAGCAGGAGAAGGAGGAGGGGGAGAAGAAGA 61

Qy    121 aatccatgaggaagtgggtcgtcgagcacaag 152
          |    | || ||||  ||     ||| | |||
Db     60 AGGAGAAGAAGAAGGAGGAGAAGGAGGAGAAG 29



Search completed: February  7, 2002, 10:59:33
Job time: 4959 sec

                          GenCore version 4.5
                Copyright (c) 1993 - 2000  Compugen Ltd.


OM nucleic - nucleic search, using sw model

Run on:          February  7, 2002, 09:10:34 ; Search time 172.96 Seconds
                                              (without alignments)
                                              551.268 Million cell updates/sec

Title:           US-09-394-745-5893
Perfect score:   421
Sequence:        1 gaaaaaaataactcggaaaa.........ccatgttggttcctgcatgc 421

Scoring table:   IDENTITY_NUC
                 Gapop 10.0 , Gapext 1.0

Searched:        351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters:      702406

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
```

Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	49.6	11.8	7218	1	US-08-232-463-14	Sequence 14, Appl
	2	38.2	9.1	289	4	US-09-007-005-17	Sequence 17, Appl
	3	38.2	9.1	289	4	US-09-244-796-17	Sequence 17, Appl
	4	34	8.1	3756	2	US-08-576-626A-1	Sequence 1, Appli
c	5	33.2	7.9	1763	6	5198542-1	Patent No. 5198542
	6	32.6	7.7	1027	2	US-08-867-087B-54	Sequence 54, Appl
	7	31.8	7.6	248	4	US-09-007-005-32	Sequence 32, Appl
	8	31.8	7.6	248	4	US-09-244-796-32	Sequence 32, Appl
	9	31.8	7.6	277	4	US-09-007-005-3	Sequence 3, Appli
	10	31.8	7.6	277	4	US-09-244-796-3	Sequence 3, Appli
c	11	31.8	7.6	2426	5	PCT-US91-09422-20	Sequence 20, Appl
c	12	31.4	7.5	24979	2	US-08-147-777-3	Sequence 3, Appli
c	13	31.4	7.5	24979	3	US-08-452-872-3	Sequence 3, Appli
c	14	31.4	7.5	24979	5	PCT-US93-03985-3	Sequence 3, Appli
c	15	30.8	7.3	4897	6	5196516-7	Patent No. 5196516
	16	30.2	7.2	1066	1	US-08-314-309A-18	Sequence 18, Appl
	17	30.2	7.2	2076	5	PCT-US91-08442-1	Sequence 1, Appli
	18	30.2	7.2	2079	3	US-08-557-210A-6	Sequence 6, Appli
	19	30.2	7.2	2109	3	US-08-557-210A-7	Sequence 7, Appli
	20	30.2	7.2	2109	3	US-08-557-210A-8	Sequence 8, Appli
	21	30.2	7.2	2300	4	US-09-344-438-2	Sequence 2, Appli
	22	30.2	7.2	2580	2	US-08-887-798-1	Sequence 1, Appli
	23	30.2	7.2	3237	2	US-08-419-075-26	Sequence 26, Appl
	24	30.2	7.2	3337	1	US-08-072-610-1	Sequence 1, Appli
	25	30.2	7.2	3337	2	US-08-719-822B-1	Sequence 1, Appli
	26	30.2	7.2	3337	4	US-09-092-458-1	Sequence 1, Appli
c	27	30.2	7.2	6623	2	US-08-244-434-36	Sequence 36, Appl
c	28	30.2	7.2	6630	2	US-08-244-434-37	Sequence 37, Appl
	29	30.2	7.2	7305	1	US-08-286-740-4	Sequence 4, Appli
	30	30.2	7.2	7305	5	PCT-US95-09576-4	Sequence 4, Appli
c	31	30.2	7.2	8575	5	PCT-US92-08258-6	Sequence 6, Appli
c	32	30.2	7.2	8932	2	US-08-252-493C-8	Sequence 8, Appli
c	33	30.2	7.2	8932	3	US-09-276-197-8	Sequence 8, Appli
c	34	30.2	7.2	10580	1	US-08-196-259-1	Sequence 1, Appli
c	35	30.2	7.2	11616	1	US-08-196-259-2	Sequence 2, Appli
	36	30	7.1	118	1	US-08-182-175A-52	Sequence 52, Appl
	37	30	7.1	118	1	US-08-474-633A-61	Sequence 61, Appl
	38	30	7.1	118	5	PCT-US92-06412-52	Sequence 52, Appl

c	39	30	7.1	1223	4	US-08-957-351-4	Sequence 4, Appli
c	40	30	7.1	1223	4	US-08-957-351-29	Sequence 29, Appl
c	41	30	7.1	1240	4	US-08-957-351-8	Sequence 8, Appli
	42	30	7.1	1276	4	US-09-177-325-2	Sequence 2, Appli
	43	30	7.1	1276	4	US-09-411-812A-2	Sequence 2, Appli
	44	30	7.1	1276	4	US-09-590-113-2	Sequence 2, Appli
	45	29.8	7.1	15144	3	US-08-458-434A-6	Sequence 6, Appli

ALIGNMENTS

```
RESULT   1
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
;  GENERAL INFORMATION:
;    APPLICANT:  DORNER, F.
;    APPLICANT:  SCHEIFLINGER, F.
;    APPLICANT:  FALKNER, F. G.
;    TITLE OF INVENTION:  RECOMBINANT FOWLPOX VIRUS
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  1800 Diagonal Road, Suite 500
;      CITY:  Alexandria
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22313-0299
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/232,463
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/935,313
;      FILING DATE:
;      APPLICATION NUMBER:  EP 91 114 300.6
;      FILING DATE:  26-AUG-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  30472/114 IMMU
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)836-9300
;      TELEFAX:  (703)683-4109
;      TELEX:  899149
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7218 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
```

```
;     IMMEDIATE SOURCE:
;       CLONE:  pTZgpt-F1s
US-08-232-463-14


  Query Match             11.8%;  Score 49.6;  DB 1;  Length 7218;
  Best Local Similarity   6.4%;  Pred. No. 5.2e-06;
  Matches   10;  Conservative  106;  Mismatches   40;  Indels    0;  Gaps    0;

Qy       1 gaaaaaaataactcggaaaagaaggagacgccgaaaattcgaaaggggaggggaaagcaa 60
           :::::::: ::   :::::::::::::: :  :::::   :::::::::::::::::: ::
Db    1210 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1151

Qy      61 agctgatggcggaggcccaggggaaagcaaagcaaatggcggaggccccgagcaagatcg 120
           ::  :: :: ::::::   ::::::::: :::: ::: :: :::::    ::: ::::  :
Db    1150 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1091

Qy     121 aatccatgaggaagtgggtcgtcgagcacaagctcc 156
           ::   : :::::::: :::  :  :| | ||||||||
Db    1090 RRRRRRRRRRRRRRRRRRRRRRRRATCGCAAGCTCC 1055


RESULT   2
US-09-007-005-17
; Sequence 17, Application US/09007005B
; Patent No. 6258558
; GENERAL INFORMATION:
;  APPLICANT: Szostak, Jack W.
;  APPLICANT: Roberts, Richard W.
;  APPLICANT: Liu, Rihe
;  TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
;  TITLE OF INVENTION: FUSIONS
;  FILE REFERENCE: 00786/350003
;  CURRENT APPLICATION NUMBER: US/09/007,005B
;  CURRENT FILING DATE: 1998-01-14
;  EARLIER APPLICATION NUMBER: 60/035,963
;  EARLIER FILING DATE: 1997-01-27
;  EARLIER APPLICATION NUMBER: 60/064,491
;  EARLIER FILING DATE: 1997-11-06
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17
;   LENGTH: 289
;   TYPE: RNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Translation template
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(289)
;   OTHER INFORMATION: n = A,T,C or G
US-09-007-005-17


  Query Match              9.1%;  Score 38.2;  DB 4;  Length 289;
  Best Local Similarity    5.9%;  Pred. No. 0.005;
  Matches   14;  Conservative  103;  Mismatches  120;  Indels    0;  Gaps    0;
```

```
Qy       4 aaaaataactcggaaaagaaggagacgccgaaaattcgaaaggggaggggaaagcaaagc 63
           :|: :::| :  : :|: :|: : : :   :|:| : |: : ::: : ::: :   : :
Db      19 rarcrurarurururarcrararururarcrararurgrnrnrsrnrnrsrnrnrsrnrn 78

Qy      64 tgatggcggaggcccaggggaaagcaaagcaaatggcggaggccccgagcaagatcgaat 123
            :: :  :: :  :  : ::: :   : : : : :  :: :  :  : ::: :  :: :
Db      79 rsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrn 138

Qy     124 ccatgaggaagtgggtcgtcgagcacaagctccgagccgtaggttgcctctggctaggtg 183
            :: : ::: : :::    :: : ::: :  :  :  :: : :      :  :   :
Db     139 rsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrn 198

Qy     184 ggatcagcagttcgatcgcctacaactggtcgcggcccaatatgaagcctagcgtca 240
           :::   :::    ::    :   |:   :: |  :   :| |  : :| ::  | |:
Db     199 rsrnrnrsrnrnrsrnrnrsrcrargrcrurgrcrgrurararcrurcrururgrcr 255


RESULT   3
US-09-244-796-17
; Sequence 17, Application US/09244796
; Patent No. 6281344
; GENERAL INFORMATION:
;  APPLICANT: Szostak, Jack W.
;  APPLICANT: Roberts, Richard W.
;  APPLICANT: Liu, Rihe
;  TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
;  TITLE OF INVENTION: FUSIONS
;  FILE REFERENCE: 00786/350007
;  CURRENT APPLICATION NUMBER: US/09/244,796
;  CURRENT FILING DATE: 1999-02-05
;  EARLIER APPLICATION NUMBER: 60/035,963
;  EARLIER FILING DATE: 1997-01-27
;  EARLIER APPLICATION NUMBER: 60/064,491
;  EARLIER FILING DATE: 1997-11-06
;  EARLIER APPLICATION NUMBER: 09/007,005
;  EARLIER FILING DATE: 1998-01-14
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17
;   LENGTH: 289
;   TYPE: RNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Translation template
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(289)
;   OTHER INFORMATION: n = A,T,C or G
US-09-244-796-17


  Query Match             9.1%;  Score 38.2;  DB 4;  Length 289;
  Best Local Similarity   5.9%;  Pred. No. 0.005;
  Matches   14;  Conservative  103;  Mismatches  120;  Indels    0;  Gaps    0;

Qy       4 aaaaataactcggaaaagaaggagacgccgaaaattcgaaaggggaggggaaagcaaagc 63
```

```
          :|: :::| :  : :|: :|: : : :   :|:| : |: : ::: : ::: :   : :
Db     19 rarcrurarururururarcrararururarcrararurgrnrnrsrnrnrsrnrnrsrnrn 78

Qy     64 tgatggcggaggcccaggggaaagcaaagcaaatggcggaggccccgagcaagatcgaat 123
           :: :  :: :  :  : ::: :   : : : : :  :: :  :  : ::: :  :: :
Db     79 rsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrn 138

Qy    124 ccatgaggaagtgggtcgtcgagcacaagctccgagccgtaggttgcctctggctaggtg 183
           :: : ::: : :::    :: : ::: :  :  :  :: : :      :  :   :
Db    139 rsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrn 198

Qy    184 ggatcagcagttcgatcgcctacaactggtcgcggcccaatatgaagcctagcgtca 240
          :::   :::    ::    :   |:   :: | :   :| | : :| ::  | |:
Db    199 rsrnrnrsrnrnrsrnrnrsrcrargrcrurgrcrgrurararcrurcrururgrcr 255


RESULT   4
US-08-576-626A-1
; Sequence 1, Application US/08576626A
; Patent No. 5998194
;  GENERAL INFORMATION:
;    APPLICANT:  Summers, R.G.
;    APPLICANT:  Katz, L.
;    APPLICANT:  Donadio, S.
;    APPLICANT:  Staver, M.J.
;    TITLE OF INVENTION:  POLYKETIDE-ASSOCIATED SUGAR
;    TITLE OF INVENTION:  BIOSYNTHESIS GENES
;    NUMBER OF SEQUENCES:  60
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Abbott Laboratories
;      STREET:  100 Abbott Park Road
;      CITY:  Abbott Park
;      STATE:  Illinois
;      COUNTRY:  USA
;      ZIP:  60064-3500
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/576,626A
;      FILING DATE:  21-DEC-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Dianne Casuto
;      REGISTRATION NUMBER:  P-40,943
;      REFERENCE/DOCKET NUMBER:  5857.US.01
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (847) 938-3137
;      TELEFAX:  (847) 938-2623
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
```

```
;       LENGTH:  3756 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
US-08-576-626A-1


  Query Match              8.1%;  Score 34;  DB 2;  Length 3756;
  Best Local Similarity   53.8%;  Pred. No. 0.35;
  Matches   70;  Conservative    0;  Mismatches   60;  Indels    0;  Gaps    0;

Qy      51 gggaaagcaaagctgatggcggaggcccaggggaaagcaaagcaaatggcggaggccccg 110
           |||   ||    |||  || ||    ||||| |   || || ||  | || |||||| ||
Db    3613 GGGTGCGCCGGCCTGGAGGAGGCCAACCAGGAGCTGGCGAACCAGCTCGCCGAGGCCGCG 3672

Qy     111 agcaagatcgaatccatgaggaagtgggtcgtcgagcacaagctccgagccgtaggttgc 170
            | |  | |||   |     |  |   | ||||| |  |  |||| | ||    |  | |
Db    3673 GGGATCAGCGAGGGCGACGAGGTGCTCGACGTCGGGTTCGGGCTCGGCGCGCAGGACTTC 3732

Qy     171 ctctggctag 180
           ||||||| |
Db    3733 TTCTGGCTCG 3742


RESULT   5
5198542-1/c
;Patent No. 5198542
;    APPLICANT: ONDA, HARUO;ARIMURA, AKIRA;KIMURA, CHIHARU
;KITADA, CHIEKO
;    TITLE OF INVENTION: DNA ENCODING A PITUITARY ADENYLATE CYCLASE
;ACTIVATING PROTEIN AND USE THEREOF
;    NUMBER OF SEQUENCES: 16
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/540,105
;      FILING DATE: 10-JUN-1990
;SEQ ID NO:1:
;      LENGTH: 1763
5198542-1


  Query Match              7.9%;  Score 33.2;  DB 6;  Length 1763;
  Best Local Similarity   50.6%;  Pred. No. 0.44;
  Matches   80;  Conservative    0;  Mismatches   78;  Indels    0;  Gaps    0;

Qy      19 aagaaggagacgccgaaaattcgaaaggggaggggaaagcaaagctgatggcggaggccc 78
           |||| ||||  | |||  |   ||  ||||  |||| |  ||      ||  |    |
Db     312 AAGACGGAGGAGGCGACGAGCTGAGTGGGGGTGGGAGACAAAGTGGCCTGAAGTCCACTG 253

Qy      79 aggggaaagcaaagcaaatggcggaggccccgagcaagatcgaatccatgaggaagtggg 138
           ||  ||||| |  |  || |   || ||   |||  |||  ||      |||  || | |
Db     252 AGAAGAAAGGAGTGATAAGGAAAGAAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 193

Qy     139 tcgtcgagcacaagctccgagccgtaggttgcctctgg 176
               |||    ||  ||  | || ||   || |||||
Db     192 AGAGAGAGAGAGAGGGCCTCGGCGAAGTCGGCGTCTGG 155
```

```
RESULT   6
US-08-867-087B-54
; Sequence 54, Application US/08867087B
; Patent No. 5990386
;  GENERAL INFORMATION:
;    APPLICANT:  An, Gynheung
;    TITLE OF INVENTION:  GENES CONTROLLING FLORAL DEVELOPMENT
;    TITLE OF INVENTION:  AND APICAL DOMINANCE IN PLANTS
;    NUMBER OF SEQUENCES:  70
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Klarquist Sparkman Campbell Leigh &
;      ADDRESSEE:  Whinston, LLP
;      STREET:  One World Trade Center
;      STREET:  121 S.W. Salmon Street
;      STREET:  Suite 1600
;      CITY:  Portland
;      STATE:  Oregon
;      COUNTRY:  United States of America
;      ZIP:  97204
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Disk, 3-1/2 inch
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  MS DOS
;      SOFTWARE:  WordPerfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/867,087B
;      FILING DATE:  June 2, 1997
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  U.S. 08/323,449
;      FILING DATE:  October 14, 1994
;      APPLICATION NUMBER:  U.S. 08/485,981
;      FILING DATE:  June 7, 1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Dow, Alan. E.
;      REGISTRATION NUMBER:  35,123
;      REFERENCE/DOCKET NUMBER:  4630-47071
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (503) 226-7391
;      TELEFAX:  (503) 228-9446
;  INFORMATION FOR SEQ ID NO:  54:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1027 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double stranded
;      TOPOLOGY:  linear
US-08-867-087B-54


  Query Match             7.7%;  Score 32.6;  DB 2;  Length 1027;
  Best Local Similarity   51.7%;  Pred. No. 0.54;
  Matches   74;  Conservative    0;  Mismatches   69;  Indels    0;  Gaps    0;

Qy      20 agaaggagacgccgaaaattcgaaaggggaggggaaagcaaagctgatggcggaggccca 79
           |||  |||     || |  ||||  |||   ||||||| | |||||| |  |   |   |
Db     103 AGAGAGAGCTAGAGAGAGATCGATGGGGCGAGGGAAAGTAGAGCTGAAGCGGATCGAGAA 162
```

```
Qy      80 ggggaaagcaaagcaaatggcggaggccccgagcaagatcgaatccatgaggaagtgggt 139
           || |     |||  || ||   ||   ||| | || ||      ||| ||||  |
Db     163 CAAGATAAGCCGGCAGGTGACGTTCGCGAAGAGGAGGAACGGGCTGCTGAAGAAGGCGTA 222

Qy     140 cgtcgagcacaagctccgagccg 162
           ||    |  |  |||| | | ||
Db     223 CGAGCTGTCCGTGCTCTGCGACG 245


RESULT   7
US-09-007-005-32
; Sequence 32, Application US/09007005B
; Patent No. 6258558
; GENERAL INFORMATION:
;  APPLICANT: Szostak, Jack W.
;  APPLICANT: Roberts, Richard W.
;  APPLICANT: Liu, Rihe
;  TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
;  TITLE OF INVENTION: FUSIONS
;  FILE REFERENCE: 00786/350003
;  CURRENT APPLICATION NUMBER: US/09/007,005B
;  CURRENT FILING DATE: 1998-01-14
;  EARLIER APPLICATION NUMBER: 60/035,963
;  EARLIER FILING DATE: 1997-01-27
;  EARLIER APPLICATION NUMBER: 60/064,491
;  EARLIER FILING DATE: 1997-11-06
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 32
;   LENGTH: 248
;   TYPE: RNA
;   ORGANISM: Homo sapiens
US-09-007-005-32


  Query Match             7.6%;  Score 31.8;  DB 4;  Length 248;
  Best Local Similarity   19.5%;  Pred. No. 0.5;
  Matches   29;  Conservative   61;  Mismatches   59;  Indels    0;  Gaps    0;

Qy       4 aaaaataactcggaaaagaaggagacgccgaaaattcgaaaggggaggggaaagcaaagc 63
           :|: : :| :  :|: : :|: :|: :|  : :|   |:|: : :|:|: :|:   : :
Db      33 rarcrararururarcrararurgrgrcrurgrarargrararcrargrarararcrurg 92

Qy      64 tgatggcggaggcccaggggaaagcaaagcaaatggcggaggccccgagcaagatcgaat 123
             :::   : :    |: :|:|:   :|:|: :::|  : :| |  : : :|:| |: ::
Db      93 rarurcrurcrurgrarargrarargrarcrcrurgrcrurgrcrgrurararararcrgru 152

Qy     124 ccatgaggaagtgggtcgtcgagcacaag 152
            |: : :|:|: : :  | |: : : :|:
Db     153 rcrgrurgrararcrargrcrurgrarar 181


RESULT   8
US-09-244-796-32
; Sequence 32, Application US/09244796
; Patent No. 6281344
; GENERAL INFORMATION:
```

```
;  APPLICANT: Szostak, Jack W.
;  APPLICANT: Roberts, Richard W.
;  APPLICANT: Liu, Rihe
;  TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
;  TITLE OF INVENTION: FUSIONS
;  FILE REFERENCE: 00786/350007
;  CURRENT APPLICATION NUMBER: US/09/244,796
;  CURRENT FILING DATE: 1999-02-05
;  EARLIER APPLICATION NUMBER: 60/035,963
;  EARLIER FILING DATE: 1997-01-27
;  EARLIER APPLICATION NUMBER: 60/064,491
;  EARLIER FILING DATE: 1997-11-06
;  EARLIER APPLICATION NUMBER: 09/007,005
;  EARLIER FILING DATE: 1998-01-14
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 32
;   LENGTH: 248
;   TYPE: RNA
;   ORGANISM: Homo sapiens
US-09-244-796-32


  Query Match             7.6%;  Score 31.8;  DB 4;  Length 248;
  Best Local Similarity   19.5%;  Pred. No. 0.5;
  Matches   29;  Conservative   61;  Mismatches   59;  Indels    0;  Gaps    0;

Qy       4 aaaaataactcggaaaagaaggagacgccgaaaattcgaaaggggagggggaaagcaaagc 63
           :|: : :| :  :|: : :|: :|: :|  : :|   |:|: : :|:|: :|:   : :
Db      33 rarcrararururarcrararurgrgrcrurgrarargrararcrargrararararcrurg 92

Qy      64 tgatggcggaggcccaggggaaagcaaagcaaatggcggaggccccgagcaagatcgaat 123
             :::   : :    |: :|:|:   :|:|: :::|  : :| |  : : :|:| |: ::
Db      93 rarurcrurcrurgrarargrarargrarcrcrurgrcrurgrcrgrurararararcrgru 152

Qy     124 ccatgaggaagtgggtcgtcgagcacaag 152
            |: : :|:|: : :  | |: : : :|:
Db     153 rcrgrurgrararcrargrcrurgrarar 181


RESULT   9
US-09-007-005-3
; Sequence 3, Application US/09007005B
; Patent No. 6258558
; GENERAL INFORMATION:
;  APPLICANT: Szostak, Jack W.
;  APPLICANT: Roberts, Richard W.
;  APPLICANT: Liu, Rihe
;  TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
;  TITLE OF INVENTION: FUSIONS
;  FILE REFERENCE: 00786/350003
;  CURRENT APPLICATION NUMBER: US/09/007,005B
;  CURRENT FILING DATE: 1998-01-14
;  EARLIER APPLICATION NUMBER: 60/035,963
;  EARLIER FILING DATE: 1997-01-27
;  EARLIER APPLICATION NUMBER: 60/064,491
;  EARLIER FILING DATE: 1997-11-06
```

```
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 277
;   TYPE: RNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Translation template
US-09-007-005-3


  Query Match              7.6%;  Score 31.8;  DB 4;  Length 277;
  Best Local Similarity   19.5%;  Pred. No. 0.53;
  Matches   29;  Conservative   61;  Mismatches   59;  Indels    0;  Gaps    0;

Qy      4 aaaaataactcggaaaagaaggagacgccgaaaattcgaaaggggaggggaaagcaaagc 63
          :|: : :| :  :|: : :|: :|: :|  : :|   |:|: : :|:|: :|:   : :
Db     33 rarcrararururarcrararurgrgrcrurgrarargrararcrargrararararcrurg 92

Qy     64 tgatggcggaggcccaggggaaagcaaagcaaatggcggaggccccgagcaagatcgaat 123
            :::   : :    |: :|:|:   :|:|: :::|  : :| |  : : :|:| |: ::
Db     93 rarurcrurcrurgrarargrarargrarcrcrurgrcrurgrcrgrurararararcrgru 152

Qy    124 ccatgaggaagtgggtcgtcgagcacaag 152
           |: : :|:|: : :  | |: : : :|:
Db    153 rcrgrurgrararcrargrcrurgrarar 181


RESULT  10
US-09-244-796-3
; Sequence 3, Application US/09244796
; Patent No. 6281344
; GENERAL INFORMATION:
;  APPLICANT: Szostak, Jack W.
;  APPLICANT: Roberts, Richard W.
;  APPLICANT: Liu, Rihe
;  TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
;  TITLE OF INVENTION: FUSIONS
;  FILE REFERENCE: 00786/350007
;  CURRENT APPLICATION NUMBER: US/09/244,796
;  CURRENT FILING DATE: 1999-02-05
;  EARLIER APPLICATION NUMBER: 60/035,963
;  EARLIER FILING DATE: 1997-01-27
;  EARLIER APPLICATION NUMBER: 60/064,491
;  EARLIER FILING DATE: 1997-11-06
;  EARLIER APPLICATION NUMBER: 09/007,005
;  EARLIER FILING DATE: 1998-01-14
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 277
;   TYPE: RNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Translation template
US-09-244-796-3
```

```
  Query Match             7.6%;  Score 31.8;  DB 4;  Length 277;
  Best Local Similarity   19.5%;  Pred. No. 0.53;
  Matches   29;  Conservative   61;  Mismatches   59;  Indels    0;  Gaps    0;

Qy       4 aaaaataactcggaaaagaaggagacgccgaaaattcgaaaggggagggggaaagcaaagc 63
           :|: : :| :  :|: : :|: :|: :|  : :|   |:|: : :|:|: :|:   : :
Db      33 rarcrararururarcrararurgrgrcrurgrarargrararcrargrararararcrurg 92

Qy      64 tgatggcggaggcccaggggaaagcaaagcaaatggcggaggccccgagcaagatcgaat 123
             :::   : :    |: :|:|:   :|:|: :::|  : :| |  : : :|:| |: ::
Db      93 rarurcrurcrurgrarargrarargrarcrcrurgrcrurgrcrgrurararararcrgru 152

Qy     124 ccatgaggaagtgggtcgtcgagcacaag 152
            |: : :|:|: : :  | |: : : :|:
Db     153 rcrgrurgrararcrargrcrurgrarar 181


RESULT  11
PCT-US91-09422-20/c
; Sequence 20, Application PC/TUS9109422
;  GENERAL INFORMATION:
;    APPLICANT:  Mulvihill, Eileen R.
;    APPLICANT:  Hagen, Frederick S.
;    APPLICANT:  Houamed, Khaled M.
;    APPLICANT:  Almers, Wolfhard
;    TITLE OF INVENTION:  G PROTEIN-COUPLED GLUTAMATE RECEPTORS
;    NUMBER OF SEQUENCES:  33
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Townsend and Townsend
;      STREET:  One Market Plaza, Steuart Street Tower
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94105-1492
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US91/09422
;      FILING DATE:  19911212
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/672,007
;      FILING DATE:  18-MAR-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/648,481
;      FILING DATE:  30-JAN-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/626,806
;      FILING DATE:  12-DEC-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Parmelee, Steven W.
;      REGISTRATION NUMBER:  31,990
;      REFERENCE/DOCKET NUMBER:  13952-6PC
```

```
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 467-9600
;      TELEFAX:  (415) 543-5043
;  INFORMATION FOR SEQ ID NO:  20:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2426 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    IMMEDIATE SOURCE:
;      CLONE:  SR13
PCT-US91-09422-20


  Query Match              7.6%;  Score 31.8;  DB 5;  Length 2426;
  Best Local Similarity   52.7%;  Pred. No. 1.4;
  Matches   69;  Conservative    0;  Mismatches   62;  Indels    0;  Gaps    0;

Qy     16 gaaaagaaggagacgccgaaaattcgaaaggggaggggaaagcaaagctgatggcggagg 75
          || |   ||| |   | ||| |    |    || |||||  |   |||| | || |||||
Db   2368 GAGATCGAGGGGGAACAGAAGAGAAAATGATGGCGGGGATGGGGCAGCTCAGGGAGGAGG 2309

Qy     76 cccaggggaaagcaaagcaaatggcggaggccccgagcaagatcgaatccatgaggaagt 135
            ||||||  ||   ||      | ||| | || |    |||  | |   ||  |||||
Db   2308 GGCAGGGGGCAGGGCAGATGGCAGGGGAAGGCCAGGAGCAGAATGGAAAAATATGGAAGA 2249

Qy    136 gggtcgtcgag 146
          || |  |  ||
Db   2248 GGATGATGTAG 2238


RESULT  12
US-08-147-777-3/c
; Sequence 3, Application US/08147777
; Patent No. 5914265
;  GENERAL INFORMATION:
;    APPLICANT:  Roop, Dennis R.
;    APPLICANT:  Rothnagel, Joseph A.
;    APPLICANT:  Greenhalgh, David A.
;    APPLICANT:  Yuspa, Stuart H.
;    TITLE OF INVENTION:  KERATIN K1 EXPRESSION VECTORS
;    TITLE OF INVENTION:  AND METHODS OF USE
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LYON & LYON
;      STREET:  611 West Sixth Street
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  U.S.A.
;      ZIP:  90017
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" Diskette, 1.44 Mb storage
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  IBM MS-DOS (Version 5.0)
;      SOFTWARE:  WordPerfect (Version 5.1)
;    CURRENT APPLICATION DATA:
```

```
;       APPLICATION NUMBER:  US/08/147,777
;       FILING DATE:
;       CLASSIFICATION:  800
;     PRIOR APPLICATION DATA:
;     PRIOR APPLICATION DATA:  including application
;     PRIOR APPLICATION DATA:  described below:            two
;       APPLICATION NUMBER:  07/876,289
;       FILING DATE:  April 30, 1992
;       APPLICATION NUMBER:  Unassigned (204/144)
;       FILING DATE:  October 29, 1993
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Warburg, Richard J.
;       REGISTRATION NUMBER:  32,327
;       REFERENCE/DOCKET NUMBER:  204/153
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (213) 489-1600
;       TELEFAX:  (213) 955-0440
;       TELEX:  67-3510
;   INFORMATION FOR SEQ ID NO:  3:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  24979 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
US-08-147-777-3


  Query Match             7.5%;  Score 31.4;  DB 2;  Length 24979;
  Best Local Similarity   54.9%;  Pred. No. 5.5;
  Matches   62;  Conservative    0;  Mismatches   51;  Indels    0;  Gaps    0;

Qy      15 ggaaaagaaggagacgccgaaaattcgaaaggggaggggaaagcaaagctgatggcggag 74
           |||  || |||||  ||    |    ||    |  | | || |  |||| || |  ||||
Db    6996 GGAGGAGGAGGAGGAGCGCGGAGGAGGAGCCAGAGGAGCAAGGAGAAGCAGAGGAGGGAG 6937

Qy      75 gcccaggggaaagcaaagcaaatggcggaggccccgagcaagatcgaatccat 127
                 | |||| |||||| | | |  ||||     ||  ||||   ||| |||
Db    6936 AAGAGGAGGAAGGCAAAGGAGAAGAAGGAGAAGAAGAAGAAGAAATAATGCAT 6884


RESULT  13
US-08-452-872-3/c
; Sequence 3, Application US/08452872
; Patent No. 6057298
;  GENERAL INFORMATION:
;     APPLICANT:  Roop, Dennis R.
;     APPLICANT:  Rothnagel, Joseph A.
;     APPLICANT:  Greenhalgh, David A.
;     APPLICANT:  Yuspa, Stuart H.
;     TITLE OF INVENTION:  KERATIN K1 EXPRESSION VECTORS
;     TITLE OF INVENTION:  AND METHODS OF USE
;     NUMBER OF SEQUENCES:  5
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  LYON & LYON
;       STREET:  611 West Sixth Street
;       CITY:  Los Angeles
```

```
;       STATE:  California
;       COUNTRY:  U.S.A.
;       ZIP:  90017
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  3.5" Diskette, 1.44 Mb storage
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  IBM MS-DOS (Version 5.0)
;       SOFTWARE:  WordPerfect (Version 5.1)
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/452,872
;       FILING DATE:
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/147,777
;       FILING DATE:
;       APPLICATION NUMBER:  07/876,289
;       FILING DATE:  April 30, 1992
;       APPLICATION NUMBER:  Unassigned (204/144)
;       FILING DATE:  October 29, 1993
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Warburg, Richard J.
;       REGISTRATION NUMBER:  32,327
;       REFERENCE/DOCKET NUMBER:  204/153
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (213) 489-1600
;       TELEFAX:  (213) 955-0440
;       TELEX:  67-3510
;   INFORMATION FOR SEQ ID NO:  3:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  24979 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
US-08-452-872-3


  Query Match             7.5%;  Score 31.4;  DB 3;  Length 24979;
  Best Local Similarity   54.9%;  Pred. No. 5.5;
  Matches   62;  Conservative    0;  Mismatches   51;  Indels    0;  Gaps    0;

Qy      15 ggaaaagaaggagacgccgaaaattcgaaaggggaggggaaagcaaagctgatggcggag 74
           |||  || |||||  ||    |    ||    |  | | || |  |||| || |  ||||
Db    6996 GGAGGAGGAGGAGGAGCGCGGAGGAGGAGCCAGAGGAGCAAGGAGAAGCAGAGGAGGGAG 6937

Qy      75 gcccaggggaaagcaaagcaaatggcggaggccccgagcaagatcgaatccat 127
                 | |||| |||||| | | |  ||||     ||  ||||   ||| |||
Db    6936 AAGAGGAGGAAGGCAAAGGAGAAGAAGGAGAAGAAGAAGAAGAAATAATGCAT 6884


RESULT  14
PCT-US93-03985-3/c
; Sequence 3, Application PC/TUS9303985
;  GENERAL INFORMATION:
;    APPLICANT:  Roop, Dennis R.
;    APPLICANT:  Rothnagel, Joseph A.
;    APPLICANT:  Greenhalgh, David A.
```

```
;    APPLICANT:  Yuspa, Stuart H.
;    TITLE OF INVENTION:  DEVELOPMENT OF A VECTOR TO TARGET GENE
;    TITLE OF INVENTION:  EXPRESSION TO THE EPIDERMIS OF TRANSGENIC ANIMALS
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fulbright & Jaworski
;      STREET:  1301 McKinney, Suite 5100
;      CITY:  Houston
;      STATE:  Texas
;      COUNTRY:  U.S.A.
;      ZIP:  77010-3095
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US93/03985
;      FILING DATE:  19930428
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Paul, Thomas D.
;      REGISTRATION NUMBER:  32,714
;      REFERENCE/DOCKET NUMBER:  D-5478
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  713/651-5325
;      TELEFAX:  713/651-5246
;      TELEX:  762829
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  24979 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
PCT-US93-03985-3


  Query Match             7.5%;  Score 31.4;  DB 5;  Length 24979;
  Best Local Similarity   54.9%;  Pred. No. 5.5;
  Matches   62;  Conservative    0;  Mismatches   51;  Indels    0;  Gaps    0;

Qy      15 ggaaaagaaggagacgccgaaaattcgaaaggggaggggaaagcaaagctgatggcggag 74
           |||  || |||||  ||    |    ||    |  | | || |  |||| || |  ||||
Db    6996 GGAGGAGGAGGAGGAGCGCGGAGGAGGAGCCAGAGGAGCAAGGAGAAGCAGAGGAGGGAG 6937

Qy      75 gcccaggggaaagcaaagcaaatggcggaggccccgagcaagatcgaatccat 127
                 | |||| |||||| | | |  ||||     ||  ||||   ||| |||
Db    6936 AAGAGGAGGAAGGCAAAGGAGAAGAAGGAGAAGAAGAAGAAGAAATAATGCAT 6884


RESULT  15
5196516-7/c
;Patent No. 5196516
;    APPLICANT: SCHREURS, CHRISTA S.;METTENLEITER, THOMAS C.
```

;SIMON, ARTUR J.;LUKAS, NOEMI;RZIHA, HANNS J.
; TITLE OF INVENTION: PSEUDORABIES VIRUS VACCINE
; NUMBER OF SEQUENCES: 8
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/383,833
; FILING DATE: 21-JUL-1989
;SEQ ID NO:7:
; LENGTH: 4897
5196516-7

Query Match 7.3%; Score 30.8; DB 6; Length 4897;
Best Local Similarity 50.7%; Pred. No. 4.1;
Matches 74; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

```
Qy      14 cggaaaagaaggagacgccgaaaattcgaaaggggagggggaaagcaaagctgatggcgga 73
           ||||   ||  | ||||  |   |   | | |||| ||||| ||  | |  |  |   |
Db    4379 CGGAGGGGAGAGGGACGGAGGGGAGAGGGACGGGGGGGGGGAGAGGGACGGAGGGGAGAGG 4320

Qy      74 ggcccaggggaaagcaaagcaaatggcggaggccccgagcaagatcgaatccatgaggaa 133
           | |  |||||| ||  | |     ||  ||||  | |||       | |     | || |
Db    4319 GACGGAGGGGAGAGGGACGGGGGGGGGGAGAGGGACGGAGGGGAGAGGGACGGGGGGGGGA 4260

Qy     134 gtgggtcgtcgagcacaagctccgag 159
           | ||| ||  | | | | |  | |||
Db    4259 GAGGGACGGAGGGGAGAGGGACGGAG 4234
```

Search completed: February 7, 2002, 10:51:34
Job time: 6060 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 7, 2002, 03:18:43 ; Search time 4942.22 Seconds
(without alignments)
915.373 Million cell updates/sec

Title: US-09-394-745-5893
Perfect score: 421
Sequence: 1 gaaaaaaataactcggaaaa.....ccatgttggttcctgcatgc 421

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estin:*
4: em_estom:*
5: em_estpl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_htc:*
10: gb_est1:*
11: gb_est2:*
12: gb_htc:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rod:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		358.2	85.1	456	11	BF317968	BF317968 OV1_10_B0
c 2		297.6	70.7	508	10	AW424866	AW424866 660039G12
c 3		282.8	67.2	561	10	AW453226	AW453226 660033E05
4		248.6	59.0	396	10	BE429089	BE429089 MTD014.C0
5		237.6	56.4	649	10	AU162606	AU162606 AU162606
6		228.8	54.3	534	10	AU056970	AU056970 AU056970
7		219.8	52.2	517	10	AU070966	AU070966 AU070966
8		217.2	51.6	574	10	AW352644	AW352644 660033E05
9		212.2	50.4	464	11	D49054	D49054 RICS15677A
10		211	50.1	431	11	D22742	D22742 RICC1155A R
11		193.4	45.9	422	11	C97246	C97246 C97246 Rice
12		167.6	39.8	231	11	BF317828	BF317828 OV1_10_B0
c 13		165.4	39.3	615	10	AI967165	AI967165 614006G06
14		165.2	39.2	242	11	BF203045	BF203045 WHE1768_G
15		163	38.7	352	11	C97312	C97312 C97312 Rice
16		162	38.5	398	11	C27791	C27791 C27791 Rice
17		156	37.1	379	11	D22974	D22974 RICC1913A R
18		147.6	35.1	459	11	C28034	C28034 C28034 Rice
19		147.4	35.0	473	11	BG157171	BG157171 sab23d06.
20		145.4	34.5	502	11	BF596766	BF596766 su62d07.y
21		145.2	34.5	334	10	AI988618	AI988618 sd05d06.y
22		145.2	34.5	440	11	BG882885	BG882885 sae78h07.

	23	145.2	34.5	458	10	AW734667	AW734667 sk97h06.y
	24	145.2	34.5	464	11	BE804609	BE804609 sr84b03.y
c	25	145.2	34.5	680	10	AW348125	AW348125 GM210001A
c	26	142.4	33.8	544	10	AI711660	AI711660 605058H02
	27	142.4	33.8	559	10	AW746168	AW746168 WS1_39_B0
c	28	142.4	33.8	569	11	BG266936	BG266936 1000109A0
	29	142.4	33.8	590	10	BE360454	BE360454 DG1_63_G1
c	30	142.4	33.8	600	10	AI673899	AI673899 605038G11
	31	142.4	33.8	621	10	BE360398	BE360398 DG1_63_G1
	32	142.4	33.8	636	10	AW565110	AW565110 LG1_323_C
	33	139.8	33.2	479	10	BE445786	BE445786 WHE1453_G
c	34	139.8	33.2	505	11	BG312631	BG312631 WHE2456_A
	35	138.8	33.0	473	11	BF473251	BF473251 WHE0926_C
	36	136.6	32.4	861	10	BE558473	BE558473 HV_CEb001
	37	136.6	32.4	941	10	BE454416	BE454416 HVSMEh009
	38	135.8	32.3	448	11	T18288	T18288 5c06b10-t7
c	39	135.4	32.2	671	11	BG874201	BG874201 MEST47-E0
	40	135	32.1	276	11	D43130	D43130 D43130 Rice
	41	135	32.1	503	11	BF626192	BF626192 HVSMEa001
	42	135	32.1	521	10	AL505327	AL505327 AL505327
	43	135	32.1	557	10	AL504875	AL504875 AL504875
	44	134.4	31.9	887	11	BG365949	BG365949 HVSMEi000
	45	133.4	31.7	621	11	BF473588	BF473588 WHE0930_B

ALIGNMENTS

RESULT 1
BF317968
LOCUS BF317968 456 bp mRNA EST 21-NOV-2000
DEFINITION OV1_10_B03.b1_A002 Ovary 1 (OV1) Sorghum bicolor cDNA, mRNA sequence.
ACCESSION BF317968
VERSION BF317968.1 GI:11266505
KEYWORDS EST.
SOURCE sorghum.
ORGANISM Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE 1 (bases 1 to 456)
AUTHORS Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt ,L.H.
TITLE An EST database from Sorghum: ovaries of varying immature stages
JOURNAL Unpublished (2000)
COMMENT Contact: Cordonnier-Pratt MM
Department of Botany
The University of Georgia
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 542 1805
Email: mmpratt@uga.edu
Sequences have been trimmed to exclude PolyA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20.
Seq primer: JEN REV

```
            High quality sequence stop: 423
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .456
                     /organism="Sorghum bicolor"
                     /db_xref="taxon:4558"
                     /clone_lib="Ovary 1 (OV1)"
                     /note="Organ: Mix of ovaries of varying immature stages
                     from 8-week-old plants; Vector: pBluescript II from Lambda
                     Zap II; Site_1: XhoI; Site_2: EcoRI; The library was made
                     from poly-A RNA in the cloning vector lambda ZAP II.
                     Clones to be sequenced were prepared by mass excision."
BASE COUNT      138 a    103 c    125 g     89 t      1 others
ORIGIN


  Query Match             85.1%;  Score 358.2;  DB 11;  Length 456;
  Best Local Similarity   94.0%;  Pred. No. 2.5e-80;
  Matches  394;  Conservative    0;  Mismatches   23;  Indels    2;  Gaps    2;

Qy       4 aaaaataactcggaaaagaaggagacgccgaaaattcgaaaggggaggggaaagcaaagc 63
           |||||||||||| |||||||| |||||||||||||||||| ||| ||||||||||||||
Db       2 AAAAATAACTCCGAAAAGAAAGAGACGCCGAAAATTCGAAGGGGAAGGGGAAAGCAAAGC 61

Qy      64 tgatggcggaggcccaggggaaagcaaagcaaatggcggaggccccgagcaagatcgaat 123
             |||||||||||| |||| |||||||||||||||||||| ||||||||||||||||
Db      62 AAATGGCGGAGGCCGAGGGAAAAGCAAAGCAAATGGCGGAGGGCCCGAGCAAGATCGAAT 121

Qy     124 ccatg-aggaagtgggtcgtcgagcacaagctccgagccgtaggttgcctctggctaggt 182
           ||||| |  |||||||||||||| ||||||||||||||||||||||||||||||||||||
Db     122 CCATGAACTAAGTGGGTCGTCGACCACAAGCTCCGAGCCGTAGGTTGCCTCTGGCTAGGT 181

Qy     183 gggatcagcagttcgatcgcctacaactggtcgcggcccaatatgaagcctagcgtcaag 242
           |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db     182 GGGATCAGCAGTTCGATCGCCTACAACTGGTCGCGGCCCAATATGAAGACTAGCGTCAAG 241

Qy     243 atcatccacgcaaggttgcatgctcaagctctaaccctggctgcattagttggttctgca 302
           |||||||||||||||||||| |||| |||||||| |||||||||||||||||||||
Db     242 ATCATCCACGCAAGGTTGCATGCACAAGGTCTAACCCTAGCTGCATTAGTTGGTTCTGCA 301

Qy     303 tgcgtggagtactatgaccagaagtatggttcttctgggccaaaggtggacaaatacaca 362
           |||||||||||||||||  | ||||||||||||||||||||||||||||||||||||||
Db     302 TGCGTGGAGTACTATGACAATAAGTATGGTTCTTCTGGGCCAAAGGTGGACAAATACACA 361

Qy     363 agccaatacctggcccattcccataaagattaaaggtcgccatgttggttcctgcatgc 421
           |||||||||||||||||  | |||||||||| ||| || |||||||||||||||||  |
Db     362 AGCCAATACCTGGCCCATGCGCATAAAGATT-AAGATCTCCATGTTGGTTCCTGCATAC 419


RESULT   2
AW424866/c
LOCUS       AW424866      508 bp    mRNA            EST       09-FEB-2000
DEFINITION  660039G12.x1 660 - Mixed stages of anther and pollen Zea mays cDNA,
            mRNA sequence.
ACCESSION   AW424866
VERSION     AW424866.1  GI:6952798
KEYWORDS    EST.
```

```
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 508)
  AUTHORS   Walbot,V.
  TITLE     Maize ESTs from various cDNA libraries sequenced at Stanford
            University
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 660039  row: G  column: 12.
FEATURES             Location/Qualifiers
     source          1. .508
                     /organism="Zea mays"
                     /cultivar="Ohio43"
                     /db_xref="taxon:4577"
                     /clone_lib="660 - Mixed stages of anther and pollen"
                     /tissue_type="whole premieotic anthers to pollen shed"
                     /dev_stage="premieotic anthers to pollen shed"
                     /lab_host="XLOLR"
                     /note="Organ: anthers; Vector: Lambda Zap; Site_1: EcoRI;
                     Site_2: XhoI; Anther and pollen cDNA library.
                     Directionally sequenced with 5' end at the EcoRI site.
                     Created by Amie Franklin."
BASE COUNT      118 a    140 c    107 g    141 t      2 others
ORIGIN

  Query Match             70.7%;  Score 297.6;  DB 10;  Length 508;
  Best Local Similarity   88.8%;  Pred. No. 4.8e-65;
  Matches  332;  Conservative    0;  Mismatches   41;  Indels    1;  Gaps    1;

Qy      42 aaaggggaggggaaagcaaagctgatggcggaggcccaggggaaagcaaagcaaatggcg 101
           | ||| || |   ||     |  |  ||  ||||   || ||  |   ||||||||||||
Db     506 AGAGGAGACGCCGAAAATTCGAAGGGGGAAGAGGAGAAGAGGGGAAAGAAGCAAATGGCG 447

Qy     102 gaggccccgagcaagatcgaatccatgaggaagtgggtcgtcgagcacaagctccgagcc 161
           ||| ||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db     446 GAGTCCCCGAGCAAGATCGAATCCATGAGGAAGTGGGTCGTCGAGCCCAAGCTCCGAGCC 387

Qy     162 gtaggttgcctctggctaggtgggatcagcagttcgatcgcctacaactggtcgcggccc 221
           |||||||||||||||||||||||||||||||||||||||||| | || ||||||||||||||
Db     386 GTAGGTTGCCTCTGGCTAGGTGGGATCAGCAGTTCGATCGCATNCANCTGGTCGCGGCCC 327

Qy     222 aatatgaagcctagcgtcaagatcatccacgcaaggttgcatgctcaagctctaaccctg 281
           |||||||| |||||||||||||||||||||||||||||||| || ||||||||||||
Db     326 AATATGAAGACTAGCGTCAAGATCATCCACGCAAGGTTGCATGCGCAGGCTCTAACCCTA 267

Qy     282 gctgcattagttggttctgcatgcgtggagtactatgaccagaagtatggttcttctggg 341
           |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
```

```
Db    266 GCTGCATTAGTTGGTTCTGCATGCGTGGAGTACTACGACCAGAAGTATGGTTCTTCTGGG 207

Qy    342 ccaaaggtggacaaatacacaagccaatacctggcccattcccataaagattaaaggtcg 401
          ||||||||||||||| |||||||||||||| |||||||||| |||||||||| ||||||
Db    206 CCAAAGGTGGACAAGTACACAAGCCAATATCTGGCCCATTCGCATAAAGATT-AAGGTCC 148

Qy    402 ccatgttggttcct 415
          |||||||||||| |
Db    147 CCATGTTGGTTCAT 134


RESULT   3
AW453226/c
LOCUS       AW453226      561 bp    mRNA            EST       17-FEB-2000
DEFINITION  660033E05.y1 660 - Mixed stages of anther and pollen Zea mays cDNA,
            mRNA sequence.
ACCESSION   AW453226
VERSION     AW453226.1  GI:6994012
KEYWORDS    EST.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 561)
  AUTHORS   Walbot,V.
  TITLE     Maize ESTs from various cDNA libraries sequenced at Stanford
            University
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 660033  row: E  column: 05.
FEATURES             Location/Qualifiers
     source          1. .561
                     /organism="Zea mays"
                     /cultivar="Ohio43"
                     /db_xref="taxon:4577"
                     /clone_lib="660 - Mixed stages of anther and pollen"
                     /tissue_type="whole premieotic anthers to pollen shed"
                     /dev_stage="premieotic anthers to pollen shed"
                     /lab_host="XLOLR"
                     /note="Organ: anthers; Vector: Lambda Zap; Site_1: EcoRI;
                     Site_2: XhoI; Anther and pollen cDNA library.
                     Directionally sequenced with 5' end at the EcoRI site.
                     Created by Amie Franklin."
BASE COUNT      160 a    132 c    115 g    154 t
ORIGIN


  Query Match             67.2%;  Score 282.8;  DB 10;  Length 561;
  Best Local Similarity   95.9%;  Pred. No. 2.6e-61;
  Matches  301;  Conservative    0;  Mismatches   12;  Indels    1;  Gaps    1;
```

```
Qy     102 gaggccccgagcaagatcgaatccatgaggaagtgggtcgtcgagcacaagctccgagcc 161
           ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     561 GAGTCCCCGAGCAAGATCGAATCCATGAGGAAGTGGGTCGTCGAGCACAAGCTCCGAGCC 502

Qy     162 gtaggttgcctctggctaggtgggatcagcagttcgatcgcctacaactggtcgcggccc 221
           |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db     501 GTAGGTTGCCTCTGGCTAGGTGGGATCAGCAGTTCGATCGCATACAACTGGTCGCGGCCC 442

Qy     222 aatatgaagcctagcgtcaagatcatccacgcaaggttgcatgctcaagctctaaccctg 281
           ||||||||| ||||||||||||||||||||||||||||||||| || ||||||||||||
Db     441 AATATGAAGACTAGCGTCAAGATCATCCACGCAAGGTTGCATGCGCAGGCTCTAACCCTA 382

Qy     282 gctgcattagttggttctgcatgcgtggagtactatgaccagaagtatggttcttctggg 341
           |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db     381 GCTGCATTAGTTGGTTCTGCATGCGTGGAGTACTACGACCAGAAGTATGGTTCTTCTGGG 322

Qy     342 ccaaaggtggacaaatacacaagccaatacctggcccattcccataaagattaaaggtcg 401
           |||||||||||||| |||||||||||||| |||||||||| |||||||||| ||||||
Db     321 CCAAAGGTGGACAAGTACACAAGCCAATATCTGGCCCATTCGCATAAAGATT-AAGGTCC 263

Qy     402 ccatgttggttcct 415
           |||||||||||| |
Db     262 CCATGTTGGTTCAT 249
```

```
RESULT   4
BE429089
LOCUS       BE429089      396 bp    mRNA            EST       26-JUL-2000
DEFINITION  MTD014.C06F990624 ITEC MTD Durum Wheat Root Library Triticum
            turgidum subsp. durum cDNA clone MTD014.C06, mRNA sequence.
ACCESSION   BE429089
VERSION     BE429089.1  GI:9426932
KEYWORDS    EST.
SOURCE      durum wheat.
  ORGANISM  Triticum turgidum subsp. durum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 396)
  AUTHORS   Anderson,O.A., Appels,R., Bailey,P., Blake,T., Close,T., Cloutier
            ,S., Dubcovsky,J., Feuillet,C., Gale,M., Graner,A., Gustafson,P.,
            Herrmann,R.G., Holton,T., Jacquemin,J.M., Jia,J., Joudrier,P.,
            Langridge,P., Lazo,G.R., Lin,J.J., McGuire,P., Ogihara,Y.,
            Pecchioni,N., Qualset,C., Schuch,W., Selvaraj,G., Shariflou,M.,
            Sorrells,M., Warburton,M. and Wenzel,G.
  TITLE     International Triticeae EST Cooperative (ITEC): Production of
            Expressed Sequence Tags for Species of the Triticeae
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Joudrier P
            INRA, Unite de Biochimie et Biologie Moleculaire des Cereales
            2, place VIALA, 34060 Montpellier cedex 01 FRANCE
            Tel: 33 4 99 61 23 84
            Fax: 33 4 99 61 23 48
            Email: joudrier@ensam.inra.fr
            International Triticeae EST Cooperative (ITEC)
            http://wheat.pw.usda.gov/genome.
```

```
FEATURES             Location/Qualifiers
     source          1. .396
                     /organism="Triticum turgidum subsp. durum"
                     /cultivar="Siliana"
                     /db_xref="taxon:4567"
                     /clone="MTD014.C06"
                     /clone_lib="ITEC MTD Durum Wheat Root Library"
                     /tissue_type="root"
                     /dev_stage="3-day-old seedling, water-stressed"
                     /note="Vector: pSPORT1; T7 primers used.  See pSPORT1
                     polylinker site.  0.3-2.0 Kbp average insert size."
BASE COUNT      113 a     98 c    111 g     74 t
ORIGIN


  Query Match             59.0%;  Score 248.6;  DB 10;  Length 396;
  Best Local Similarity   81.8%;  Pred. No. 1.1e-52;
  Matches  287;  Conservative    0;  Mismatches   64;  Indels    0;  Gaps    0;

Qy      69 gcggaggcccaggggaaagcaaagcaaatggcggaggccccgagcaagatcgaatccatg 128
           || | ||     | |   || |||||||||||||||||||| ||| ||||||||||||||
Db      46 GCAGCGGAAACAGAGGCGGCGAAGCAAATGGCGGAGGCCCCAAGCCAGATCGAATCCATG 105

Qy     129 aggaagtgggtcgtcgagcacaagctccgagccgtaggttgcctctggctaggtgggatc 188
           |||||||||| ||||| |||||||||||||||||||||||||| |||||  | ||||||
Db     106 CGGAAGTGGGTGGTCGATCACAAGCTCCGAGCCGTAGGTTGCCTGTGGCTTAGCGGGATC 165

Qy     189 agcagttcgatcgcctacaactggtcgcggcccaatatgaagcctagcgtcaagatcatc 248
             ||| || ||||| |||||||||||||||||||| |||||| | ||||||||| |||||
Db     166 TCCAGCTCCATCGCGTACAACTGGTCGCGGCCCAACATGAAGACCAGCGTCAAGCTCATC 225

Qy     249 cacgcaaggttgcatgctcaagctctaaccctggctgcattagttggttctgcatgcgtg 308
           |||||||||||||||| |||||||||| | ||||| |||| | ||| |||||  ||
Db     226 CACGCAAGGTTGCATGCGCAAGCTCTAACGATCGCTGCCTTAGGTAGTTGTGCATTAGTA 285

Qy     309 gagtactatgaccagaagtatggttcttctgggccaaaggtggacaaatacacaagccaa 368
           |||||||||| ||||| || |||||||| || |||||||||||||||| ||||| ||
Db     286 GAGTACTATGAACAGAACTACGGTTCTTCAGGACCAAAGGTGGACAAATATACAAGGCAT 345

Qy     369 tacctggcccattcccataaagattaaaggtcgccatgttggttcctgcat 419
           ||| || | || || ||||||||||||   ||  |||| | | | ||| ||
Db     346 TACATGTCACACTCGCATAAAGATTAATTATCTGCATGGTCGGTGCTGGAT 396


RESULT   5
AU162606
LOCUS       AU162606      649 bp    mRNA            EST       26-OCT-2000
DEFINITION  AU162606 Rice cDNA from young root Oryza sativa cDNA clone R10541,
            mRNA sequence.
ACCESSION   AU162606
VERSION     AU162606.1  GI:11026005
KEYWORDS    EST.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
```

```
REFERENCE   1  (bases 1 to 649)
  AUTHORS   Sasaki,T. and Yamamoto,K.
  TITLE     Rice cDNA from young root (2000)
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Takuji Sasaki
            National Institute of Agrobiological Resources
            Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki
            305-8602, Japan
            Tel: 81-298-38-7441
            Fax: 81-298-38-7468
            Email: tsasaki@abr.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/
            PROJECT ='RGP'.
            R10541_2Z.
FEATURES             Location/Qualifiers
     source          1. .649
                     /organism="Oryza sativa"
                     /strain="Nipponbare"
                     /db_xref="taxon:4530"
                     /clone="R10541"
                     /clone_lib="Rice cDNA from young root"
                     /tissue_type="young root"
BASE COUNT      213 a    124 c    159 g    150 t      3 others
ORIGIN


  Query Match             56.4%;  Score 237.6;  DB 10;  Length 649;
  Best Local Similarity   81.1%;  Pred. No. 6.8e-50;
  Matches  287;  Conservative    0;  Mismatches   66;  Indels    1;  Gaps    1;

Qy      42 aaaggggaggggaaagcaaagctgatggcggaggcccaggggaaagcaaagcaaatggcg 101
           |||| | || |  |   |  |  || |  |  ||   | | |   || ||||||||||||
Db      18 AAAGAGAAGCGAGAGAAATCGGAGAGGAAGATGGGGGAAGAGGCGGCGAAGCAAATGGCG 77

Qy     102 gaggccccgagcaagatcgaatccatgaggaagtgggtcgtcgagcacaagctccgagcc 161
           ||  ||||| ||||||| ||||||||||||||||||||| |||| |||||  |||| |||
Db      78 GAANCCCCGGGCAAGATTGAATCCATGAGGAAGTGGGTCATCGACCACAA-NTCCGCGCC 136

Qy     162 gtaggttgcctctggctaggtgggatcagcagttcgatcgcctacaactggtcgcggccc 221
           |||||||||||| |||||   |||||||||||| ||||| || |||||||||||| |||||
Db     137 GTAGGTTGCCTATGGCTTACTGGGATCAGCAGCTCGATTGCGTACAACTGGTCGAGGCCC 196

Qy     222 aatatgaagcctagcgtcaagatcatccacgcaaggttgcatgctcaagctctaaccctg 281
           ||||||||| |||||||||||||||||| |||||||||||||||||| ||||| ||
Db     197 AATATGAAGACTAGCGTCAAGATCATCCATGCAAGGTTGCATGCTCAAGCCCTAACACTA 256

Qy     282 gctgcattagttggttctgcatgcgtggagtactatgaccagaagtatggttcttctggg 341
           || || ||||| || ||||||   || ||||||||||||  ||||||  ||  | |||||
Db     257 GCAGCGTTAGTGGGATCTGCAATGGTAGAGTACTATGACGCGAAGTACGGCACATCTGGA 316

Qy     342 ccaaaggtggacaaatacacaagccaatacctggcccattcccataaagattaa 395
           || |||||||||| |||||||||||||||||||| ||||| ||||||||||||
Db     317 CCGAAGGTGGACAAGTACACAAGCCAATACCTGGCGCATTCACATAAAGATTAA 370


RESULT   6
AU056970
LOCUS       AU056970      534 bp    mRNA            EST       29-APR-1999
```

```
DEFINITION  AU056970 Oryza sativa mature leaf Nipponbare Oryza sativa cDNA
            clone S21027_1A, mRNA sequence.
ACCESSION   AU056970
VERSION     AU056970.1  GI:4715854
KEYWORDS    EST.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 534)
  AUTHORS   Yamamoto,K. and Sasaki,T.
  TITLE     Rice cDNA from mature leaf
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Takuji Sasaki
            National Institute of Agrobiological Resources
            Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki
            305-8602, Japan
            Tel: 81-298-38-7441
            Fax: 81-298-38-7468
            Email: tsasaki@abr.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/
            PROJECT ='RGP'.
FEATURES             Location/Qualifiers
     source          1. .534
                     /organism="Oryza sativa"
                     /strain="Nipponbare"
                     /db_xref="taxon:4530"
                     /clone="S21027_1A"
                     /clone_lib="Oryza sativa mature leaf Nipponbare"
                     /tissue_type="mature leaf"
BASE COUNT      167 a    109 c    149 g    109 t
ORIGIN
```

```
  Query Match             54.3%;  Score 228.8;  DB 10;  Length 534;
  Best Local Similarity   74.5%;  Pred. No. 1.1e-47;
  Matches  301;  Conservative    0;  Mismatches  102;  Indels    1;  Gaps    1;

Qy      17 aaaagaaggagacgccgaaaattcgaaaggggagggaaagcaaagctgatggcggaggc 76
           |||  ||| |||     |||   || || | || | || || ||       ||  || |
Db      24 AAAGAAAGAAGAAAAACAAAGCGCGTAAAGAGAAGCGAGAGAAATCGGAGAGGAAGATGG 83

Qy      77 ccaggggaaagcaaagcaaatggcggaggccccgagcaagatcgaatccatgaggaagtg 136
               |    |   | |||||||||||| |||||| ||||||| ||||||||||||||||
Db      84 GGGAAGAGGCGGCGAACAAATGGCGGAAGCCCCGGGCAAGATTGAATCCATGAGGAAGTG 143

Qy     137 ggtcgtcgagcacaagctccgagccgtaggttgcctctggctaggtgggatcagcagttc 196
           |||| |||| |||||||||||| |||||| ||||||| |||||   |||||||||||| ||
Db     144 GGTCATCGACCACAAGCTCCGCGCCGTA-GTTGCCTATGGCTTACTGGGATCAGCAGCTC 202

Qy     197 gatcgcctacaactggtcgcggcccaatatgaagcctagcgtcaagatcatccacgcaag 256
           ||| || |||||||||||| |||||||||||||| |||||||||||||||||| |||||
Db     203 GATTGCGTACAACTGGTCGAGGCCCAATATGAAGACTAGCGTCAAGATCATCCATGCAAG 262

Qy     257 gttgcatgctcaagctctaaccctggctgcattagttggttctgcatgcgtggagtacta 316
           |||||||||||||| ||||| || || || ||||| || ||||||   || ||||||||
Db     263 GTTGCATGCTCAAGCCCTAACACTAGCAGCGTTAGTGGGATCTGCAATGGTAGAGTACTA 322
```

```
Qy    317 tgaccagaagtatggttcttctgggccaaaggtggacaaatacacaagccaatacctggc 376
          ||||  |||||| ||  | ||||| || |||  |||||| ||||||||||||||||||||
Db    323 TGACGCGAAGTACGGCACATCTGGACCGAAGTGGGACAAGTACACAAGCCAATACCTGGG 382

Qy    377 ccattcccataaagattaaaggtcgccatgttggttcctgcatg 420
          |   | |    ||   | ||| ||  ||| || | |  || |||
Db    383 CGCATTCACATAAAGATTAAGATCTTCATATTCGCTGTTGGATG 426
```

```
RESULT   7
AU070966
LOCUS       AU070966      517 bp    mRNA            EST       10-JUN-1999
DEFINITION  AU070966 Rice cDNA from young root Oryza sativa cDNA clone
            R10541_1A, mRNA sequence.
ACCESSION   AU070966
VERSION     AU070966.1  GI:5038856
KEYWORDS    EST.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 517)
  AUTHORS   Yamamoto,K. and Sasaki,T.
  TITLE     Rice cDNA from young root
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Takuji Sasaki
            National Institute of Agrobiological Resources
            Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki
            305-8602, Japan
            Tel: 81-298-38-7441
            Fax: 81-298-38-7468
            Email: tsasaki@abr.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/
            PROJECT ='RGP'.
FEATURES             Location/Qualifiers
     source          1. .517
                     /organism="Oryza sativa"
                     /strain="Nipponbare"
                     /db_xref="taxon:4530"
                     /clone="R10541_1A"
                     /clone_lib="Rice cDNA from young root"
                     /tissue_type="young root"
BASE COUNT      165 a    103 c    142 g    102 t      5 others
ORIGIN
```

```
  Query Match             52.2%;  Score 219.8;  DB 10;  Length 517;
  Best Local Similarity   76.8%;  Pred. No. 2e-45;
  Matches  291;  Conservative    0;  Mismatches   86;  Indels    2;  Gaps    2;

Qy     17 aaaagaaggagacgccgaaaattcgaaaggggaggggaaagcaaagctgatggcggaggc 76
          |||  ||| |||     |||   || || | || | || || ||       ||  || |
Db     28 AAAGAAAGAAGAAAAACAAAGCGCGTAAAGAGAAGCGAGAGAAATCGGAGAGGAAGATGG 87

Qy     77 ccaggggaaagcaaagcaaatggcggaggccccgagcaagatcgaatccatgaggaagtg 136
                |    |   | ||||| ||||| | |||| ||||||| ||||||||||||||||
```

```
Db      88 GGGAAGAGGCGGCGAACAAATNGCGGAAG-CCCGGGCAAGATTGAATCCATGAGGAAGTG 146

Qy     137 ggtcgtcgagcacaagctccgagccgtaggttgcctctggctaggtgggatcagcagttc 196
           |||| |||| |||||||||||| |||||| ||||||| |||||   |||||||||||| ||
Db     147 GGTCATCGACCACAAGCTCCGCGCCGTA-GTTGCCTATGGCTTACTGGGATCAGCAGCTC 205

Qy     197 gatcgcctacaactggtcgcggcccaatatgaagcctagcgtcaagatcatccacgcaag 256
           ||| || |||||||||||| |||||||||||||| | |||||||||||||||| |||||
Db     206 GATTGCGTACAACTGGTCGAGGCCCAATATGAAGACNAGCGTCAAGATCATCCATGCAAG 265

Qy     257 gttgcatgctcaagctctaaccctggctgcattagttggttctgcatgcgtggagtacta 316
           |||||||||||||||| ||||| |  || || ||||| || || |||   || ||||||||
Db     266 GTTGCATGCTCAAGCCCTAACACNAGCAGCGTTAGTGGGATCNGCAATGGTAGAGTACTA 325

Qy     317 tgaccagaagtatggttcttctgggccaaaggtggacaaatacacaagccaatacctggc 376
           ||||  |||||| ||  | ||||| || ||| .|||||| ||||||||||||||||||||
Db     326 TGACGCGAAGTACGGCACATCTGGACCGAAGTGGGACAAGTACACAAGCCAATACCTGGC 385

Qy     377 ccattcccataaagattaa 395
            ||||| ||||||||||||
Db     386 GCATTCACATAAAGATTAA 404
```

```
RESULT   8
AW352644
LOCUS       AW352644      574 bp    mRNA            EST       02-FEB-2000
DEFINITION  660033E05.x1 660 - Mixed stages of anther and pollen Zea mays cDNA,
            mRNA sequence.
ACCESSION   AW352644
VERSION     AW352644.1  GI:6851634
KEYWORDS    EST.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 574)
  AUTHORS   Walbot,V.
  TITLE     Maize ESTs from various cDNA libraries sequenced at Stanford
            University
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 660033  row: E  column: 05.
FEATURES             Location/Qualifiers
     source          1. .574
                     /organism="Zea mays"
                     /cultivar="Ohio43"
                     /db_xref="taxon:4577"
                     /clone_lib="660 - Mixed stages of anther and pollen"
                     /tissue_type="whole premieotic anthers to pollen shed"
                     /dev_stage="premieotic anthers to pollen shed"
```

```
                     /lab_host="XLOLR"
                     /note="Organ: anthers; Vector: Lambda Zap; Site_1: EcoRI;
                     Site_2: XhoI; Anther and pollen cDNA library.
                     Directionally sequenced with 5' end at the EcoRI site.
                     Created by Amie Franklin."
BASE COUNT      173 a    132 c    151 g    118 t
ORIGIN


  Query Match             51.6%;  Score 217.2;  DB 10;  Length 574;
  Best Local Similarity   92.7%;  Pred. No. 9.3e-45;
  Matches  228;  Conservative    0;  Mismatches   18;  Indels    0;  Gaps    0;

Qy      68 ggcggaggcccaggggaaagcaaagcaaatggcggaggccccgagcaagatcgaatccat 127
           ||  ||||   || ||  |  ||||||||||||||| ||||||||||||||||||||||
Db     329 GGAAGAGGAGAAGAGGGGAAAGAAGCAAATGGCGGAGTCCCCGAGCAAGATCGAATCCAT 388

Qy     128 gaggaagtgggtcgtcgagcacaagctccgagccgtaggttgcctctggctaggtgggat 187
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     389 GAGGAAGTGGGTCGTCGAGCACAAGCTCCGAGCCGTAGGTTGCCTCTGGCTAGGTGGGAT 448

Qy     188 cagcagttcgatcgcctacaactggtcgcggcccaatatgaagcctagcgtcaagatcat 247
           |||||||||||||| |||||||||||||||||||||||| ||||||||||||||||
Db     449 CAGCAGTTCGATCGCATACAACTGGTCGCGGCCCAATATGAAGACTAGCGTCAAGATCAT 508

Qy     248 ccacgcaaggttgcatgctcaagctctaaccctggctgcattagttggttctgcatgcgt 307
           ||||||| |||||||||| || |||||||||| |||||||||||||||||||||||||||
Db     509 CCACGCAGGGTTGCATGCGCAGGCTCTAACCCTAGCTGCATTAGTTGGTTCTGCATGCGT 568

Qy     308 ggagta 313
           ||||||
Db     569 GGAGTA 574


RESULT   9
D49054
LOCUS       D49054         464 bp    mRNA            EST       02-AUG-1995
DEFINITION  RICS15677A Rice green shoot Oryza sativa cDNA, mRNA sequence.
ACCESSION   D49054
VERSION     D49054.1  GI:702763
KEYWORDS    EST.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 464)
  AUTHORS   Sasaki,T., Miyao,A. and Yamamoto,K.
  TITLE     Rice cDNA from callus 1995
  JOURNAL   Unpublished (1995)
COMMENT     Contact: Takuji Sasaki
            National Institute of Agrobiological Resources
            Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki
            305-8602, Japan
            Tel: 81-298-38-7441
            Fax: 81-298-38-7468
            Email: tsasaki@abr.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/.
```

```
FEATURES             Location/Qualifiers
     source          1. .464
                     /organism="Oryza sativa"
                     /strain="Nipponbare"
                     /db_xref="taxon:4530"
                     /clone_lib="Rice green shoot"
                     /note="Green shoot (8 days old)"
BASE COUNT      146 a     90 c    136 g     87 t      5 others
ORIGIN


  Query Match             50.4%;  Score 212.2;  DB 11;  Length 464;
  Best Local Similarity   73.1%;  Pred. No. 1.6e-43;
  Matches  296;  Conservative    0;  Mismatches  107;  Indels    2;  Gaps    2;

Qy      17 aaaagaaggagacgccgaaaattcgaaaggggagggggaaagcaaagctgatggcggaggc 76
           |||  ||| |||     |||   || || | || | || || ||       ||  || |
Db      32 AAAGAAAGAAGAAAAACAAAGCGCGTAAAGAGAAGCGAGAGAAATCGGAGAGGAAGATGG 91

Qy      77 ccaggggaaagcaaagcaaatggcggaggccccgagcaagatcgaatccatgaggaagtg 136
                |    |   |  ||   |  |    |||| ||||||| ||||||||||||||||
Db      92 GGGAAGAGGCGGCGAANAANATGGCGGAAGCCCGGGCAAGATTGAATCCATGAGGAAGTG 151

Qy     137 ggtcgtcgagcacaagctccgagccgtaggttgcctctggctaggtgggatcagcagttc 196
           |||| |||| |||||||||| |||||| ||||||| |||||   |||||||||||  ||
Db     152 GGTCATCGACCACAAGCTCCGCGCCGTA-GTTGCCTATGGCTTACTGGGATCAGCAGCTC 210

Qy     197 gatcgcctacaactggtcgcggcccaatatgaagcctagcgtcaagatcatccacgcaag 256
           ||| || |||||||||||| |||||||||||||| |||||||||||||||||| |||||
Db     211 GATTGCGTACAACTGGTCGAGGCCCAATATGAAGACTAGCGTCAAGATCATCCATGCAAG 270

Qy     257 gttgcatgctcaagctctaaccctggctgcattagttggttctgcatgcgtggagtacta 316
           |||||||||||||| ||||| || || || ||||| || ||||||   || ||||||||
Db     271 GTTGCATGCTCAAGCCCTAACACTAGCAGCGTTAGTGGGATCTGCAATGGTAGAGTACTA 330

Qy     317 tgaccagaagta-tggttcttctgggccaaaggtggacaaatacacaagccaatacctgg 375
           ||||  ||||||  ||  | ||||| || |||  |||||| ||||||||||||||||||
Db     331 TGACGCGAAGTACGGGCACATCTGGACCNAAGTGGGACAAGTACACAAGCCAATACCTGG 390

Qy     376 cccattcccataaagattaaaggtcgccatgttggttcctgcatg 420
             ||||| ||| ||| |||| |     ||| || | |  ||  ||
Db     391 NGCATTCACATTAAGGTTAAGGTCTTTCATATTCGCTGTTGGGTG 435


RESULT  10
D22742
LOCUS       D22742         431 bp    mRNA            EST       08-JUL-1999
DEFINITION  RICC1155A Rice callus Oryza sativa cDNA clone C1155_1A, mRNA
            sequence.
ACCESSION   D22742
VERSION     D22742.1  GI:431806
KEYWORDS    EST.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
```

```
REFERENCE   1  (bases 1 to 431)
  AUTHORS   Sasaki,T. and Minobe,Y.
  TITLE     Rice cDNA from callus
  JOURNAL   Unpublished (1994)
COMMENT     Contact: Takuji Sasaki
            National Institute of Agrobiological Resources
            Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki
            305-8602, Japan
            Tel: 81-298-38-7441
            Fax: 81-298-38-7468
            Email: tsasaki@abr.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/
            PROJECT ='RGP'.
FEATURES             Location/Qualifiers
     source          1. .431
                     /organism="Oryza sativa"
                     /strain="cultivar Nipponbare, sub_species Japonica"
                     /db_xref="taxon:4530"
                     /clone="C1155_1A"
                     /clone_lib="Rice callus"
                     /note="Vector: pBluescript II SK+; Site_1: SalI; Site_2:
                     NotI; cDNA prepared from rice callus mRNAs by using
                     oligo(dT) as a primer and ligating to the SalI-NotI site
                     of pBluescript II SK+ phagemid.  "
BASE COUNT      144 a     86 c    128 g     73 t
ORIGIN


  Query Match             50.1%;  Score 211;  DB 11;  Length 431;
  Best Local Similarity   75.2%;  Pred. No. 3.3e-43;
  Matches  318;  Conservative    0;  Mismatches   95;  Indels   10;  Gaps    4;

Qy      2 aaaaaaataactcggaaaagaaggagacgccgaaaattcgaaaggggaggggaaagcaaa 61
          || |||| | |||| |||| ||    | |  |||||    |  | | |  |  |||| |
Db      9 AAGAAAAGAGCTCGAAAAAAAAAAGAAAGAAGAAAAACAAAGCGCGTAAAGAGAAGCGAG 68

Qy     62 gctgatggcggaggcccaggggaaa-------gcaaagcaaatggcggaggccccgagca 114
              || |  ||||   | |||  |       || |||||||||||||| | |||| |||
Db     69 AGAAATCGGAGAGGAAGATGGGGGAAGAGGCGGCGAAGCAAATGGCGGAAG-CCCGGGCA 127

Qy    115 agatcgaatccatgaggaagtgggtcgtcgagcacaagctccgagccgtaggttgcctct 174
          |||| |||||||||||||||||||| |||| |||||||||| |||||| ||||||| |
Db    128 AGATTGAATCCATGAGGAAGTGGGTCATCGACCACAAGCTCCGCGCCGTA-GTTGCCTAT 186

Qy    175 ggctaggtgggatcagcagttcgatcgcctacaactggtcgcggcccaatatgaagccta 234
          ||||   |||||||||||| ||||| || |||||||||||| |||||||||||||| |||
Db    187 GGCTTACTGGGATCAGCAGCTCGATTGCGTACAACTGGTCGAGGCCCAATATGAAGACTA 246

Qy    235 gcgtcaagatcatccacgcaaggttgcatgctcaagctctaaccctggctgcattagt-t 293
          ||||||||||||||| |||||||||||||||||| ||||| || || || |||||
Db    247 GCGTCAAGATCATCCATGCAAGGTTGCATGCTCAAGCCCTAACACTAGCAGCGTTAGTGG 306

Qy    294 ggttctgcatgcgtggagtactatgaccagaagtatggttcttctgggccaaaggtggac 353
          || ||||||   || ||||||||||| ||||||  ||  | |||||  || | |  ||||
Db    307 GGATCTGCAATGGTAGAGTACTATGACGCGAAGTACGGCACATCTGGACCGAGGTGGGAC 366

Qy    354 aaatacacaagccaatacctggcccattcccataaagattaaaggtcgccatgttggttc 413
          || |||||||||||||||||| ||||| ||||| ||||||      |    | ||||
```

```
Db    367 AAGTACACAAGCCAATACCTGGCGCATTCACATAAGGATTAAGATCTTCATATTCGGTTG 426

Qy    414 ctg 416
          ||
Db    427 TTG 429
```

RESULT 11
C97246

LOCUS	C97246	422 bp	mRNA	EST	19-OCT-1998

DEFINITION C97246 Rice callus Oryza sativa cDNA clone C53369_1A, mRNA sequence.
ACCESSION C97246
VERSION C97246.1 GI:3759988
KEYWORDS EST.
SOURCE Oryza sativa.
ORGANISM Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.
REFERENCE 1 (bases 1 to 422)
AUTHORS Sasaki,T. and Minobe,Y.
TITLE Rice cDNA from callus
JOURNAL Unpublished (1994)
COMMENT Contact: Takuji Sasaki
National Institute of Agrobiological Resources
Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki 305-8602, Japan
Tel: 81-298-38-7441
Fax: 81-298-38-7468
Email: tsasaki@abr.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/
PROJECT ='RGP'.
FEATURES Location/Qualifiers
source 1. .422
/organism="Oryza sativa"
/strain="cultivar Nipponbare, sub_species Japonica"
/db_xref="taxon:4530"
/clone="C53369_1A"
/clone_lib="Rice callus"
/note="Vector: pBluescript II SK+; Site_1: SalI; Site_2: NotI; cDNA prepared from rice callus mRNAs by using oligo(dT) as a primer and ligating to the SalI-NotI site of pBluescript II SK+ phagemid. "
BASE COUNT 141 a 87 c 116 g 75 t 3 others
ORIGIN

Query Match 45.9%; Score 193.4; DB 11; Length 422;
Best Local Similarity 74.4%; Pred. No. 9e-39;
Matches 296; Conservative 0; Mismatches 94; Indels 8; Gaps 4;

```
Qy      5 aaaataactcggaaaagaaggagacgccgaaaattcgaaaggggaggggaaagcaaagct 64
          |||| | |||| |||| ||    | |  |||||    |  | | |  |  |||| |
Db      2 AAAAGAGCTCGAAAAAAAAAAGAAAGAAGAAAAACAAAGCGCGTAAAGAGAAGCGAGAGA 61

Qy     65 gatggcggaggcccagggg----aaagcaaagcaaatggcggaggccccgagcaagatcg 120
           || |  ||||   | |||     | ||   |  ||  |  |   |||   ||||||| |
```

```
Db      62 AATCGGAGAGGAAGATGGGGGAAGAGGCGGCGACAAATGGCGGAACCCGNGGCAAGATTG 121

Qy     121 aatccatgaggaagtgggtcgtcgagcacaagctccgagccgtaggttgcctctggctag 180
           |||||||||||||||||||| |||| |||||||||| |||||| ||||||| |||
Db     122 AATCCATGAGGAAGTGGGTCATCGACCACAAGCTCCGCGCCGTA-GTTGCCTATGGTNTA 180

Qy     181 gtgggatcagcagttcgatcgcctacaactggtcgcggcccaatatgaagcctagcgtca 240
            |||||||||||| ||||| || |||||||||||| |||||||||||||| |||||||||
Db     181 CTGGGATCAGCAGCTCGATTGCGTACAACTGGTCGAGGCCCAATATGAAGACTAGCGTCA 240

Qy     241 agatcatccacgcaaggttgcatgctcaagctctaaccctggctgcattagt-tggttct 299
           |||||||||| ||||||||||||||||||| ||||| || || || |||||  || |||
Db     241 AGATCATCCATGCAAGGTTGCATGCTCAAGCCCTAACACTAGCAGCGTTAGTGGGGATCT 300

Qy     300 gcatgcgtggagtactatgaccagaagtatggttcttctgggccaaaggtggac--aaat 357
           |||   || ||||||||||||  |||||| ||  | ||| | || |||  ||||  || |
Db     301 GCAATGGTAGAGTACTATGACGCGAAGTACGGCACATCTTGACCCAAGTNGGACCAAAGT 360

Qy     358 acacaagccaatacctggcccattcccataaagattaa 395
           |||||||||||||||||| ||||| ||||||||||||
Db     361 ACACAAGCCAATACCTGGCGCATTCACATAAAGATTAA 398
```

```
RESULT  12
BF317828
LOCUS       BF317828      231 bp    mRNA            EST       21-NOV-2000
DEFINITION  OV1_10_B02.g1_A002 Ovary 1 (OV1) Sorghum bicolor cDNA, mRNA
            sequence.
ACCESSION   BF317828
VERSION     BF317828.1  GI:11266334
KEYWORDS    EST.
SOURCE      sorghum.
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 231)
  AUTHORS   Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt
            ,L.H.
  TITLE     An EST database from Sorghum: ovaries of varying immature stages
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Cordonnier-Pratt MM
            Department of Botany
            The University of Georgia
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 542 1805
            Email: mmpratt@uga.edu
            Sequences have been trimmed to exclude PolyA, vector and regions
            below Phred quality 16. The threshold for highest quality sequence
            is 20.
            Seq primer: PolyTMix
            High quality sequence start: 69
            High quality sequence stop: 200
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .231
```

```
                     /organism="Sorghum bicolor"
                     /db_xref="taxon:4558"
                     /clone_lib="Ovary 1 (OV1)"
                     /note="Organ: Mix of ovaries of varying immature stages
                     from 8-week-old plants; Vector: pBluescript II from Lambda
                     Zap II; Site_1: XhoI; Site_2: EcoRI; The library was made
                     from poly-A RNA in the cloning vector lambda ZAP II.
                     Clones to be sequenced were prepared by mass excision."
BASE COUNT       65 a     52 c     54 g     60 t
ORIGIN
```

Query Match 39.8%; Score 167.6; DB 11; Length 231;
Best Local Similarity 94.8%; Pred. No. 2.7e-32;
Matches 184; Conservative 0; Mismatches 9; Indels 1; Gaps 1;

```
Qy    225 atgaagcctagcgtcaagatcatccacgcaaggttgcatgctcaagctctaaccctggct 284
          ||||||||||||||||||||||||||||||||||||||||||| |||| ||||||||| |||
Db      1 ATGAAGCCTAGCGTCAAGATCATCCACGCAAGGTTGCATGCACAAGGTCTAACCCTAGCT 60

Qy    285 gcattagttggttctgcatgcgtggagtactatgaccagaagtatggttcttctgggcca 344
          |||||||||||||||||||||||||||||||||||| | ||||||||||||||||||||
Db     61 GCATTAGTTGGTTCTGCATGCGTGGAGTACTATGACAATAAGTATGGTTCTTCTGGGCCA 120

Qy    345 aaggtggacaaatacacaagccaatacctggcccattcccataaagattaaaggtcgcca 404
          |||||||||||||||||||||||||||||||||||| | |||||||||| ||| || |||
Db    121 AAGGTGGACAAATACACAAGCCAATACCTGGCCCATGCGCATAAAGATT-AAGATCTCCA 179

Qy    405 tgttggttcctgca 418
          ||||||||||||||
Db    180 TGTTGGTTCCTGCA 193
```

```
RESULT  13
AI967165/c
LOCUS       AI967165      615 bp    mRNA            EST       23-AUG-1999
DEFINITION  614006G06.x3 614 - root cDNA library from Walbot Lab Zea mays cDNA,
            mRNA sequence.
ACCESSION   AI967165
VERSION     AI967165.1  GI:5762117
KEYWORDS    EST.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 615)
  AUTHORS   Walbot,V.
  TITLE     Maize ESTs from various cDNA libraries sequenced at Stanford
            University
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
```

```
            Email: walbot@stanford.edu
            Plate: 614006  row: G  column: 06.
FEATURES             Location/Qualifiers
     source          1. .615
                     /organism="Zea mays"
                     /cultivar="W23"
                     /db_xref="taxon:4577"
                     /clone_lib="614 - root cDNA library from Walbot Lab"
                     /tissue_type="root"
                     /dev_stage="3-4 days old"
                     /lab_host="XLOLR"
                     /note="Organ: root; Vector: pBlueScriptII SK+; Site_1:
                     EcoRI; Site_2: XhoI; 3-4 days old root tissue from Walbot
                     Lab (LM)"
BASE COUNT      210 a    133 c    117 g    154 t      1 others
ORIGIN


  Query Match             39.3%;  Score 165.4;  DB 10;  Length 615;
  Best Local Similarity   99.4%;  Pred. No. 1.1e-31;
  Matches  166;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy     255 aggttgcatgctcaagctctaaccctggctgcattagttggttctgcatgcgtggagtac 314
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     491 AGGTTGCATGCTCAAGCTCTAACCCTGGCTGCATTAGTTGGTTCTGCATGCGTGGAGTAC 432

Qy     315 tatgaccagaagtatggttcttctgggccaaaggtggacaaatacacaagccaatacctg 374
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     431 TATGACCAGAAGTATGGTTCTTCTGGGCCAAAGGTGGACAAATACACAAGCCAATACCTG 372

Qy     375 gcccattcccataaagattaaaggtcgccatgttggttcctgcatgc 421
           |||||||| ||||||||||||||||||||||||||||||||||||
Db     371 GCCCATTCGCATAAAGATTAAAGGTCGCCATGTTGGTTCCTGCATGC 325


RESULT  14
BF203045
LOCUS       BF203045      242 bp    mRNA            EST       06-NOV-2000
DEFINITION  WHE1768_G09_M18ZS Wheat pre-anthesis spike cDNA library Triticum
            aestivum cDNA clone WHE1768_G09_M18, mRNA sequence.
ACCESSION   BF203045
VERSION     BF203045.1  GI:11117799
KEYWORDS    EST.
SOURCE      bread wheat.
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 242)
  AUTHORS   Anderson,O.D., Chao,S., Choi,D.W., Close,T.J., Fenton,R.D., Han
            ,P.S., Hsia,C.C., Kang,Y., Lazo,G.R., Miller,R., Rausch,C.J.,
            Seaton,C.L. and Tong,J.C.
  TITLE     The structure and function of the expressed portion of the wheat
            genomes - Pre-anthesis spike cDNA library
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Olin Anderson
            US Department of Agriculture, Agriculture Research Service, Pacific
```

West Area, Western Regional Research Center
800 Buchanan Street, Albany, CA 94710, USA
Tel: 5105595773
Fax: 5105595818
Email: oandersn@pw.usda.gov
Sequence have been trimmed to remove vector sequence and low quality sequence with phred score less than 20
Seq primer: Stratagene SK primer.

```
FEATURES             Location/Qualifiers
     source          1. .242
                     /organism="Triticum aestivum"
                     /cultivar="Chinese Spring"
                     /db_xref="taxon:4565"
                     /clone="WHE1768_G09_M18"
                     /clone_lib="Wheat pre-anthesis spike cDNA library"
                     /tissue_type="Spike before anthesis"
                     /dev_stage="Adult plant"
                     /lab_host="E. coli SOLR"
                     /note="Vector: Lambda Uni-ZAP XR, excised phagemid;
                     Site_1: EcoRI; Site_2: XhoI; Plants were grown in the
                     greenhouse. Whole spike with awns trimmed, white, green
                     and yellow anther were collected and total RNA, and
                     poly(A) RNA were prepared, a cDNA library was made, and
                     the cDNA clones were in vivo excised to give pBluescript
                     phagemids in the TJ Close lab (Choi, Close, Fenton) at
                     the University of California, Riverside. Plasmid DNA
                     preparations and DNA sequencing were performed in the OD
                     Anderson lab (all other authors)."
BASE COUNT       60 a     72 c     70 g     39 t      1 others
ORIGIN
```

```
  Query Match             39.2%;  Score 165.2;  DB 11;  Length 242;
  Best Local Similarity   88.2%;  Pred. No. 1.1e-31;
  Matches  179;  Conservative    0;  Mismatches   24;  Indels    0;  Gaps    0;

Qy      87 gcaaagcaaatggcggaggccccgagcaagatcgaatccatgaggaagtgggtcgtcgag 146
           || |||||||||||||||||||||||||||||||||||||||| |||||||||| |||||
Db      40 GCGAAGCAAATGGCGGAGGCCCCGAGCAAGATCGAATCCATGCGGAAGTGGGTGGTCGAT 99

Qy     147 cacaagctccgagccgtaggttgcctctggctaggtgggatcagcagttcgatcgcctac 206
           |||||||||||||||||||||||||| |||||    ||||||  ||| || ||||| |||
Db     100 CACAAGCTCCGAGCCGTAGGTTGCCTGTGGCTTANCGGGATCTCCAGCTCCATCGCGTAC 159

Qy     207 aactggtcgcggcccaatatgaagcctagcgtcaagatcatccacgcaaggttgcatgct 266
           |||||||||||||||| |||||| | ||||||||| |||||||||||||||||||| ||
Db     160 AACTGGTCGCGGCCCAACATGAAGACCAGCGTCAAGCTCATCCACGCAAGGTTGCACGCG 219

Qy     267 caagctctaaccctggctgcatt 289
           |||||||||||  | ||||| ||
Db     220 CAAGCTCTAACGATCGCTGCCTT 242
```

```
RESULT  15
C97312
LOCUS       C97312         352 bp    mRNA            EST       19-OCT-1998
DEFINITION  C97312 Rice callus Oryza sativa cDNA clone C53986_1A, mRNA
```

```
            sequence.
ACCESSION   C97312
VERSION     C97312.1  GI:3760054
KEYWORDS    EST.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 352)
  AUTHORS   Sasaki,T. and Minobe,Y.
  TITLE     Rice cDNA from callus
  JOURNAL   Unpublished (1994)
COMMENT     Contact: Takuji Sasaki
            National Institute of Agrobiological Resources
            Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki
            305-8602, Japan
            Tel: 81-298-38-7441
            Fax: 81-298-38-7468
            Email: tsasaki@abr.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/
            PROJECT ='RGP'.
FEATURES             Location/Qualifiers
     source          1. .352
                     /organism="Oryza sativa"
                     /strain="cultivar Nipponbare, sub_species Japonica"
                     /db_xref="taxon:4530"
                     /clone="C53986_1A"
                     /clone_lib="Rice callus"
                     /note="Vector: pBluescript II SK+; Site_1: SalI; Site_2:
                     NotI; cDNA prepared from rice callus mRNAs by using
                     oligo(dT) as a primer and ligating to the SalI-NotI site
                     of pBluescript II SK+ phagemid.  "
BASE COUNT       78 a     88 c    120 g     65 t      1 others
ORIGIN


  Query Match             38.7%;  Score 163;  DB 11;  Length 352;
  Best Local Similarity   75.7%;  Pred. No. 4.1e-31;
  Matches  202;  Conservative    0;  Mismatches   65;  Indels    0;  Gaps    0;

Qy      68 ggcggaggcccaggggaaagcaaagcaaatggcggaggccccgagcaagatcgaatccat 127
           || ||||  |  |||     |  |  ||||||||||||    ||||| |||  | || ||
Db      43 GGAGGAGAGCTTGGGTTGGCCTGACGAAATGGCGGAGGAGAAGAGCACGATGCAGTCGAT 102

Qy     128 gaggaagtgggtcgtcgagcacaagctccgagccgtaggttgcctctggctaggtgggat 187
           |||| |||||||||||||| |||||||||| ||||| ||    || |||||   |||| |
Db     103 GAGGGAGTGGGTCGTCGACCACAAGCTCCGGGCCGTCGGGACGCTGTGGCTGACTGGGGT 162

Qy     188 cagcagttcgatcgcctacaactggtcgcggcccaatatgaagcctagcgtcaagatcat 247
               || |||||||| |||||||||||| |||||   |||||| | ||||||||||||||
Db     163 GGCGAGCTCGATCGCGTACAACTGGTCGAGGCCCGGCATGAAGACCAGCGTCAAGATCAT 222

Qy     248 ccacgcaaggttgcatgctcaagctctaaccctggctgcattagttggttctgcatgcgt 307
           |||||| |||||||||||||| || || || |||||||||| | ||| ||||||   ||
Db     223 CCACGCCAGGTTGCATGCTCAGGCCCTCACACTGGCTGCATTGGCTGGCTCTGCACTGGT 282

Qy     308 ggagtactatgaccagaagtatggttc 334
```

```
          ||||||||| |||||   ||  |||||
Db    283 GGAGTACTACGACCATCGGTCAGGTTC 309
```

Search completed: February 7, 2002, 08:20:37
Job time: 18114 sec